

QY	481	ggaggaaaaaagaaataagcaaaaacaaacccgctttctcggactcggccttatcccat	540
Db	481	GGAGGAAAAAGAAATAGCAAAAACAACCCCGTTTCTGGACTGGCGCTTATCCCAATTTAT	540
QY	541	tggagcaaatgcacctgtgcgcggagagatgaaagaataatctactgcttgytanaaagacac	600
Db	541	TGGACCAATGTCGGTTGCCGGAGAGATGAAGAAATATACTGCTTTGTGAAAAAGACATC	600
QY	601	tggcagggttggggtgtgtctcccaagaagatgycacgtcttcaacgftatgctttcccat	660
Db	601	TGGCAGGGTGGGGTGTGTCTCCACAAAGATGGCACTGCTCTACGCTAGCTTTTCCAT	660
QY	661	ggcccaacgttcccacatcattcgctctgtcgtgagcgacagfytgtcaagaagccgagtg	720
Db	661	GGCCCAAGGTTCCCAATCATTTCCGCTTGTGGAGGCGACAGTGTGCATGAAGCGGGG	720
QY	721	gagccttcacatgtctgcgcagtgaggagaccggtttgtatgaccaatgggatatgcgat	780
Db	721	GAGCTTACCAATGCTGGCGCAATGGGGAAACGGTTGTATGATCAACCAAGGATATCCGAT	780
QY	781	gcagagatgatctgcagcgacgtcggcctcagtggtcattgtcacaagcatcggcatcaagg	840
Db	781	GCAGAGATGATCTGCAGGCAAGCTGGCGCTCAGTGGCATTTCCAAACATGCGATCAGCA	840
QY	841	tatttggggaagggctcgtgcacgltatgttgaatgaatgactcgtcaccttggaaatgag	900
Db	841	TATTTTGGGGAAGGGCTGTGGCCCGATTATTTGGATGAAGTACGCTGACTGGGAATGAG	900
QY	901	cttccaatgtgaggttccaaagagctctcggggagagcatatcgttggccataagaa	960
Db	901	CTTCCAATGTGACAGTGTCCAAAGAGCTCTGGGGAGAGCATATCTGTGCCATTAAGAA	960
QY	961	gatcgtgagatgtccctgtaccccctcaacagatcgggtgtacagacttcaggttggaaa	1020
Db	961	GATCGTGAAGTGTCTCTGTACCCCTCTACAGATGGGGTCTACAGACTTCAGTGGGAAA	1020
QY	1021	ggcagccaatggaaggttcgcttggagaatalttacagagcgcaatcgggaaactcgtgtat	1080
Db	1021	GGCAGCCATGAGGGTCTGCTTGGAGAGTATTTACAGAGCGCAAGTGGGAACTGTCTGTAT	1080
QY	1081	gatggcttgagactcgtcgtgaatacactaagtgcttgcgacagtttgggatttaatatgt	1140
Db	1081	GATGGCTGAGACTGTCGTAATATCATAGTGTGTTGCGACAGTGGCATTTAAATATGT	1140
QY	1141	aaacaagcatctgcacaacattttgaagaagaacagggcccaatcagrttggatgaagct	1200
Db	1141	AAACAAGCATCTGCCAACCATTTTGAAGAAAGCACAGGGCCCATATGTGTGATACGTC	1200
QY	1201	agctgtctcaagaaagaaacccaagattctctcagtgcttccaagcgacagttcgggaaagcat	1260
Db	1201	AGCTGTCTCAGAAAGAAACCAAGATTCTTCAGTGTTCACAGGGACAGTGGGAAGGCAT	1260
QY	1261	gactcgcagcccccgcggaagatgttaagatctgcctgtctaccccttggcggcgaaggaacaag	1320
Db	1261	GACTGCAGCCCAACCCGGAAGATGTTAGCATTTGCTGTACCTGTGCGCGCCAGGAGACACAG	1320
QY	1321	ctctcctcggggtttctcctgtccaagctatgtatcgtgagaaataaagaagaagagcagtg	1380
Db	1321	CTCTCTCTGGGTTTCTCTGTCCAGACTGATGTGATGGAGAAAAATTAAGAAAGACAGCACTG	1380
QY	1381	gaaggtttatcaatcagtcgagtcggggaacaatctgtatgatatgatatgatatgaagat	1440
Db	1381	GAGGTTTTATCAATGCGCAGATGGGGAACAATCTGTGATGATGATGATGACATTAAGAT	1440
QY	1441	gcagctgtgatactgtgtcagcttggctacaagggctcctcggcagagcaagaacatggct	1500
Db	1441	GCAGCTGTGATCTGTCTCTCCTCAAGCTTGGCTTACAAAGGCTCTCCAGAGCAAGAACATGGCT	1500
QY	1501	tacttctggagaagaaagaaacccaatccatctgtgataatgtgaagtgcacgggaaatgag	1560
Db	1501	TACTTCTGGAGAAGAAAGAACCCATCTCATGTGTGAATATGTAAAGTGCACAGGAATATAG	1560

QY	1561	aggtctctgggtgagctcgtatcaagcaagatattggaagacaactgcgcgcacagctgaa	1620
Db	1561	AGGTCTCTGGGTGAGCTCGTATCAAGCAAGATATTGGAAGACAACTGCCGCCACAGTGA	1620
QY	1621	gatgcagaagctattcttgattattcttggaagaagagctcaagctaaacgtataaagag	1680
Db	1621	GATGCAGAAGCTATTCTTGATTATTTTGGCAGAGAGCCTCGAGTAAACGTAATTAAGAG	1680
QY	1681	tcctctcatcctctgtcttgagcttgaattactgcacgctcggcagaagcggatcattgct	1740
Db	1681	TCCTCTCATCTCGTTGTGTGGGCTTGAGATTACTGCACGCGGAGAGAGGGATCATTTGGT	1740
QY	1741	gggaaaaattcttctaaggggtgtgttgcccttggcaggttccctccgcgtcaagtcac	1800
Db	1741	GGGAAAAATTCTTTAAGGGGGGTGGCTTGCCCTTGCGACAGTTTCCCTCGCTGAAGTCA	1800
QY	1801	catgtagatgcaagcgcctcctcgcggagcctcctcctcctcctcctcctcctcctc	1860
Db	1801	CATGAGATGCGCAGGCTCTCTTGCGGGGCTACGCTCTCTAGTAGTGCTGGGCTCTCA	1860
QY	1861	gcagcacactcttccaagaggatctggaacagcaactcctcctcctcctcctcctcctc	1920
Db	1861	GCAGCACACTCTTCCAGAGGATAGGAAACAGCACTAGAGCTATGCTTAGGGTTGGA	1920
QY	1921	gattatcatctctgtgtacacagagaggttgaggaataattggcagttcaacagattgtg	1980
Db	1921	GATTATCATCTACTCTGTGTACCAAGAGAGGATTGAGGAATATGGAGTTCAACACATTGTG	1980
QY	1981	attatctggaggtctgcagccgcagccgaatgattatgatatagcctcctggttaattca	2040
Db	1981	ATTATCTGGGAGTTCGACCCGACCGAGTATTGATTAATAGATAGCCCTGGTTAGATTCA	2040
QY	2041	ggaaccagaagagcaatctgccaagattcagaacagccattgtttgcacagcctgttaacc	2100
Db	2041	GGACCAAGAAGCAATGTGCCAATTCAGAGGCAATGTTTGCCAGCCTGTTTACACTC	2100
QY	2101	tggagagagagccacagaanaacagcatccaactgttatacatacagatctgggtgacaca	2160
Db	2101	TGGAGAGAGAGCCACAGAAACAGCATCCCACTGTTACATACAGATGGGGTGGACACA	2160
QY	2161	ggacagagcctatttaagaagaacttaacaagaagccattcccttaactccataaagttc	2220
Db	2161	GGACGAGCCTATTTCMAAACACTACCAACAAGACGCAATCCCTTACTTCTTAAGAGTTT	2220
QY	2221	tgttaagaacgtttaaagggttggttcttaacaggagaatgcttfgtgcctggaacccat	2280
Db	2221	TGTGAAGACCTTTAAGGGTTCGGTTTACAGGAGAACTCTTGTGCTGGAAACTCTCAT	2280
QY	2281	gaacacaaacgctgtgacaagctctccaggagagacagcgcgagagacacatcattgtgaa	2340
Db	2281	GAACACAAACCGTGGACACTCCAGGGGAGACAGGGAGGAGCACTCATGTGTGAACGG	2340
QY	2341	cccgagagagagcttgggtgtgtgtatgggtgtgacctctctgggggtatagctgtggag	2400
Db	2341	CCCGAGAGAGACTGGGTGTGTGTGGGGGTGACCTCTGGGGGTATGGCTGTGGAGTCA	2400
QY	2401	gattccctgtgttcttaccaaagctcagcccttgaactctgtgataaagaagtgtaac	2460
Db	2401	GATTCTCTGCTGTTTATACCAAGTCTAGCCCTTGTATCTTGATTAATAAGTGTAC	2460
QY	2461	aaactgtaatctctcatgtaaaccttcaaaacagacatlttaacaaattggaacattgac	2520
Db	2461	AAACTGTAACTCTCATGAAACTTCAAGCAGCATTTTAACAAATGGAATACTTTAAC	2520
QY	2521	ccccactattagacactcagcagagatgacacaaaacggcgaag	2562
Db	2521	CCCCACTATTAGCACTCAGCAGATGACAAACAAACGGCAAG	2562

QY	841	tattttggggaagggctctggccccagttatgttgataagtaacgtctcacgggaatgag	900
Db	1043	TATTTTGGGGAAGGCTCTGCCCAAGTTATGTGATGAATACCTCTGACTGGGAATGAG	1102
QY	901	ctttcaattgagcagttgtccaaagagctccttgggaagcataactgttggccataaagaa	960
Db	1103	CTTTCAATTGAGCAGATGTCCTCAAAAGAGCTCTGGGAGAGCATAACTGTGGCCATAAAGAA	1162
QY	961	gatgtctggagttgtcctgttaacctctaacagatgaggttcatcagacttgcagttgggaaa	1020
Db	1163	GATGCTGGAGTGTCTGTATCCCTCTTAACAGATGGGTCATCAACACTTGCAGGGTGGGAAA	1222
QY	1021	ggcagccatgagaggttcgcttggaggtatatacagagccagttggggaactgtctgtgat	1080
Db	1223	GGCAGCCATGAGGGTCCCTTGGAGGTATATTACAGAGCCAGTGGGCACTGTCTGTGAT	1282
QY	1081	gatgtctggagtgaagtcgtaaatacacaagctgttgttcgcacagtttgggaattaaatagt	1140
Db	1283	GATGCTGGAGCTGAGCTGGAATACATACGTGGTTGTGCAGAGTTGGGATTTAAATATAGT	1342
QY	1141	aaacaaagcatcttgcacacatttttgaaagaagcaagggcccatagtgttggatgacgtc	1200
Db	1343	AAACAAGCATCTGCCCCAACCTTTTGAAGAAAGCAAGAGGCCATATGTTGGATGAAGTC	1402
QY	1201	agctgtctcaggaaaggaanaacacagatttcttcagttgtccagggcgacagttgggaagcat	1260
Db	1403	AGCTGCTCAGGAAGGAAGAACCAATTTCTTCAGTGTTCAGGCGACAGTGGGGAAAGCAT	1462
QY	1261	gactgacgaccccgacgaagatgtttaagatltgcctgtcactacctggcgcgagaggaacag	1320
Db	1463	GACTGACGACCGCGAGATGTTAGCATTTGCCGTGACCTGGCGGAGGGGACACAGG	1522
QY	1321	cctctctctgggttttccctgtcagacttgaatgtgagaaataaagaagaagcaggtg	1380
Db	1523	CTCTCTCTGGGTTTCTCTGCAACACTGATGATGGAATAAGAAAGAGCGAGTG	1582
QY	1381	gaggttttatacaatgacagttgggaaacatctgttatgatgtatgagttgactgtataagat	1440
Db	1583	GAGGTTTTATCATATGGCCAGTGGGACACATCTGTATGATGATGACTGATAAAGAT	1642
QY	1441	gcagctgtatctgtctgcacagcttggctaaagggctcctgcagagcaagaacacatgct	1500
Db	1643	GCAGCTGTATCTGTCTGTACGTTGGCTACAAAGGTCCTGCCAGAGCAAGAACCATGGCT	1702
QY	1501	tactttgagaaagaaagaagccatccatctgttgataatgttgaagtgacagaagaatgag	1560
Db	1703	TACTTTGGAGGAAGAAAGAACCCATCCATGCGATGATGTGAAGTCCACAGGAAATGAG	1762
QY	1561	aaggtcttggtgcgaagctgtatcaaggaagatatgtgaaagaacaaactcggccaaagtgaa	1620
Db	1766	AGGTCTTGTGGCTGACTGTAATCAAGCAAGATTTGGAAAGAACAACTCCGCCACAGTGAA	1822
QY	1621	gatgcagagatctaatgttgataatttggcaagaagccctcaagttaacagtaataaagag	1680
Db	1823	GATGCAAGAGTTATTTGTGATTTTTTGGCAAGAGCCCTCAGGTAACAGTAATAAAGAG	1882
QY	1681	tccctctcatctgttttgggtcttgaagatatactgcaacgcttcgcagaaagcagatcatgt	1740
Db	1883	TCCCTCTCATCTGTTGTGTGGCTTGAGATTACTGCACCCGTGGCAGAAAGCGGATCATTTGT	1942
QY	1741	gggaaaaaatctttaaggggtgtgttgcccttgcagagtttccctcggctgaaatcatcc	1800
Db	1943	GGGAAAAATCTTTAAGGGGTGTGTGGCTTGGAGGTTTCCCTCCGCGTGAATCTATCC	2002
QY	1801	cattgagatgacagagctcctctgcggagctacagctctctgagtagctctgtgtctaca	1860
Db	2003	CATTGAGATGGCAGGCTCCTCTCGGGGCTAACGCTCTGATATCTCTGGGTCTCTACA	2062
QY	1861	gcagcacactgttttcaagaagatataggcaaacagactagagactgtctgtttaaggttga	1920
Db	2063	GCAGCACACTGTTTCAAGAAGATATGGCAACAGCACTAGAGGCTATGCTGTTAAGGTTGGA	2122
QY	1921	gattatcatactctgttaccagagaggtttgaggaagaanaatttggagtccaacagattgtg	1980

Query Match	54.6%	Score 1398	DB 1	Length 2614
Best Local Similarity	84.2%	Pred. No. 0		
Db	2123	GATTATCATCTCGTGGTACACAGAGAGTTGTGAGAGAAATTGGAGTTCAACAGATTGG	2182	
Qy	1981	atcatcggagagatcgaccgcgcagtgatltgacatagccctggtltagatacaa	2040	
Db	2183	ATTCTATCGGGATTCGACCCCGACCCGACCTGATGTATGTGACATAGGCCCTGGTATATACAA	2242	
Qy	2041	ggaccagagagagaahtgtgcacatttcacacacacatglttttggcagccttttaccac	2100	
Db	2243	GGACCAAGAGAGCAATGTGCAATTCACACACCACATGTTTGGCCAGCCTGTTTACCACTC	2302	
Qy	2101	tggagagagagccacagaaaaacagcatccaactgtatcaatacagagatgaggtagacaa	2160	
Db	2303	TGGAGAGAGAGCCACAGAAATACAGATTCACACTTTTCAATACAGATGGGGTACACA	2362	
Qy	2161	ggacgagcctattccaagaacatacaacaagcagccattccctacttctaagaagttt	2220	
Db	2363	GGACGAGCCTATTCCAAGAACATCAACACAGCAGCCATTCCCTTACTCTTAAAGGTTT	2422	
Qy	2221	tgtaagaagacgttataaaggttgcgtttaaaggaagatccttgttgcgttgaaccccat	2280	
Db	2423	TGTGAAGACGTTATTAAGGGCTCGTTTAAAGGAGAAAGCTTTGTGCTGGAAACCTTCAT	2482	
Qy	2281	gaacacaacgcgctggagcaagctccagaggagacagcggagagaccacatcgttgaacg	2340	
Db	2483	GAACACAACCGCGTGGACACCTCCACGGAGACACGGAGACACCACTCATGTGTGAACGG	2542	
Qy	2341	cccgagagagagcttggtygtgtatgtgggtgtaacctctctgggggtatggctgtgagtc	2400	
Db	2543	CCCGAGAGAGACTGGGGTGGTGTGGGGTGACCTCTGGGGGTATGGCTGTGAGATCAAG	2602	
Qy	2401	gattctcctggtgtttataaccaaaagctcagcctttagcctttagataaaaagtgcacc	2460	
Db	2603	GATTCCTCGTGGTGTATATACCAAGTCTCAGCTTTGACCTTGGATTAATAAGTGTACCC	2662	
Qy	2461	aaactgaattcttcacatgagaaacttcaaaagcagcatttaacaaatgaaacttgaac	2520	
Db	2663	AAACTGTAATTCCTTCATGAGAAATTCGAAGCAGCTTTAAACAATATGAGAAATTTGAAAC	2722	
Qy	2521	ccccactattagcaactcagacagatgacaaacagggcaag	2562	
Db	2723	CCCCACTATTAGCACTCAGCAGAGATGACAACTATGCGAAG	2764	
RESULT	3			
ID	X19027	standard; DNA; 2614 BP.		
AC	X19027			
DT	13-MAY-1999	(first entry)		
DE	Mouse serine protease BSSP-3 encoding DNA.			
KW	Serine protease; BSSP-3; brain tissue; ss.			
OS	Mus sp.			
PN	MO9905290-ATL			
PD	UT-FEB-1999			
PR	24-JUL-1998	J03324.		
PR	24-JUL-1997	JP-213965.		
PA	(SDNRK-)-SUNFORY LTD.			
PI	Tsuruoka N, Yamaguchi N, Yamashiro K;			
DR	WPI: 99-142942/12.			
P	P-PSDB: w99088.			
PT	New serine protease expressed in brain tissue - used in screening			
PS	for potential serine protease inhibitors for drug use			
PT	Example 1; Page 45-50; 69pp; Japanese.			
CC	The present sequence encodes a serine protease designated BSSP-3, which			
CC	is isolated from mouse brain tissue. Transformants may be used to			
CC	produce the enzyme or its partial sequences. Products from the present			
CC	invention are used for screening for potential peptide or non-peptide			
CC	serine protease inhibitors or expression regulators for use as drugs.			
SO	Sequence 2614 BP; 614 A; 671 C; 799 G; 530 T;			

FT CDS 24..2309
 FT /tag- b
 FT sig_peptide 24..86
 FT /tag- c
 FT mat_peptide 87..2306
 FT /tag- d
 FT misc_feature 90..275
 FT /tag- e
 FT /note- "proline rich, basic segment"
 FT misc_feature 276..494
 FT /tag- f
 FT /note- "Kringles domain"
 FT misc_feature 519..824
 FT /tag- g
 FT /note- "SRCR domain 1"
 FT misc_feature 840..1142
 FT /tag- h
 FT /note- "SRCR domain 2"
 FT misc_feature 1179..1484
 FT /tag- i
 FT /note- "SRCR domain 3"
 FT misc_feature 1536..2306
 FT /tag- j
 FT /note- "proteolytic domain"
 FT misc_feature 1707..1709
 FT /tag- k
 FT /note- "histidine of the catalytic triade"
 FT misc_feature 1857..1859
 FT /tag- l
 FT /note- "aspartic acid of the catalytic triade"
 FT misc_feature 2154..2156
 FT /tag- m
 FT /note- "serine of the catalytic triade"
 FT 3'UTR 2307..2356
 FT /tag- n
 FT polyA_signal 2324..2329
 FT /tag- o
 FT polyA_signal 2331..2336
 FT /tag- p
 FT polyA_signal 2357..2376
 FT /tag- q
 PN M09849322-A1.
 PD 05-NOV-1998
 PF 24-APR-1998..1B0625..
 PR 26-APR-1997: CH-000966
 PA (SOMD//) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 P-PSDB: W83362.
 CC New human and murine neurotrophin - used, e.g. for inhibiting
 CC gene therapy and in drug development
 CC Claim 1: Page 29-32: 50pp: English.
 CC The present sequence encodes mouse neurotrophin. Neurotrophin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 CC Sequence 2376 BP: 589 A: 586 C: 708 G: 493 T:

Query Match 54.4%; Score 1393.2; DB 1; Length 2376;
 Best Local Similarity 84.5%; Pred. No. 0;
 Matches 1566; Conservative 0; Mismatches 288; Indels 0; Gaps 0;

QY 618 tctcagaagatgagcagctgctcagctgagcttccctccagccagcttccat 677
 DB 458 TGCCCAAGGCAAGTACAGCTGGGCTACTGCTGCTTGTGCTCAAGGCCGCGCTGCCCT 517
 QY 678 cattgccttctgtagagcagcagctgtagatgaagccgggtgagctcaccatgctg 737
 DB 518 CATTGCGCTTGTGTGGGGAACAGTGGGATGATGAGTCAAGTGAAGCTGACAGCTGG 577
 QY 738 ccaagtggggaacccgttctgtagatgacccaatggatgtagtccgtagcaagaatgctcag 797
 DB 578 CCAAGTGGGGAACATCTGTGACGACCAATGGCAATGAGCAATGAGCATGATCTGTAG 637
 QY 798 gcaagctggagctcagctgagctcagcagcagcagcagcagcagcagcagcagcagcagc 857
 DB 638 GCAAGCTGGGCTAGTGGCTGCTTCCAAAGCATGCTGCTGCTGCTGCTGCTGCTGCTG 697
 QY 858 tggcccaagttagttagatgaagtagcagctcagctcagcagcagcagcagcagcagcagc 917
 DB 698 TGGCCCAATATGTGTGAGTGAAGTACGTGCAACCGAAGAGCTGTCAATTGAGCAATG 757
 QY 918 tccaagagctcctggggagagacatactgtgccaataagaagatgctgagctgctg 977
 DB 758 TCCAAGAGTCTCTGGGGGGAACATACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
 QY 978 taacctcacaagatgagctgagctcagcagcagcagcagcagcagcagcagcagcagcagc 1037
 DB 818 TGTCTCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 877
 QY 1038 cttagagatataaagaagccagctgagcagcagcagcagcagcagcagcagcagcagcagc 1097
 DB 878 CTTGAGAGTCTACTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 937
 QY 1098 gaatacagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1157
 DB 938 GAACACATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 997
 QY 1158 ccaatttgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1217
 DB 998 CCAATTTGATGAGCAGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1057
 QY 1218 aaccagattcttcaatgctcagcagcagcagcagcagcagcagcagcagcagcagcagc 1277
 DB 1058 AGTACAGCTTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1117
 QY 1278 agatttagcattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1337
 DB 1118 AGATGTGGGGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1177
 QY 1338 tgcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1397
 DB 1178 CATCAGACTAGTGTGATGAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
 QY 1398 ccaagtggggaacaatctgtagatgtagatgtagatgtagatgtagatgtagatgtagatg 1457
 DB 1238 CCAATGGGGAACAATCTGGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATG 1297
 QY 1458 tcaagtgtgctacaagagctgctgcaagcagcagcagcagcagcagcagcagcagcagcagc 1517
 DB 1298 GCAACTTGGCTATTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1357
 QY 1518 agagccacacacagctgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtag 1577
 DB 1358 AGGCCCATCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1417
 QY 1578 tatcaagcaagatattgtagaagacaacacacacacacacacacacacacacacacacacac 1637
 DB 1418 TGTCAACAACAAGATGAGAGGACACACACACACACACACACACACACACACACACACAC 1477
 QY 1638 tgattatttggcaagagcctcagctgtagaagtaataaagaagctccctcattctgttgg 1697

Db 2662 ACCTTCCAGTCCAGCCAGACCCAGACACTTGGCCAACTTCTGTCATCAC 2721
 Oy 1332 -----tttccctgcagactgagatgagaaataagaagaag 1373
 Db 2722 ACAGAGATCGATTCACACTTGGCCCTGAGACGTGTAATGAGAGGTGACAGGTGTCAGG 2781
 Oy 1374 acgagatgagaggttttatacaatgagcagtgagaaacatctgtagatgagtgactga 1433
 Db 2782 CCGAGTGGAGGTCCTTACCAAGGCTCGGGCACCGTGTGTATGATACACAGGAGAC 2841
 Oy 1434 taagagatgagcctgtagctgtagcagctgtagctgtagcagcagcagcagcagcagc 1493
 Db 2842 CATATGATGCCACAGTGTGTCGAGGACAGCTGGCTGGGCTGAGTGTGACAGCCACAG 2901
 Oy 1494 catgcttacttctgagagaaagaaacacatcattgtagataatgtagaagtcgacagc 1553
 Db 2902 AATGCGCCAGTTTGGCCAGGCTCAGGACCAATGTCCTGATGATGATGCGCTGCTCAGG 2961
 Oy 1554 aatatgagagctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1613
 Db 1962 ACACGAGTCTTACCTGTGAGAGCTGCCCCCAATGCTGCTGCTGCCCAACTGTGCGCA 3021
 Oy 1614 cagtgaagatgcagagatgattg 1637
 Db 3022 TCATGATGATGCTGCTGCTGCTGCTG 3045

RESULT 6
 ID 004926
 AC 004926
 DT 23-OCT-1990 (first entry)
 DE Sequence encoding portion of scavenger receptor protein with
 DE affinity for acylated low density lipoprotein (LDL).
 KW Acylated low density lipoprotein; LDL; atherosclerotic plaque;
 FH Key
 FT cds
 Location/Qualifiers
 cds 4.1362
 /*tag= a

FN WO9005748-A
 PD 31-MAY-1990.
 PF 14-NOV-1989; 005116.
 PR 15-NOV-1988; US-272002.
 PA (MAST) Massachusetts Inst Tech.
 PI Engelienner J;
 DR WPI: 90-193408/25.
 P-PSDB: R05509.
 PT New receptor protein -
 PT has affinity for acetylated low density lipoprotein and
 PT corresponding antibodies and DNA sequences.
 CC Disclosure: 1sh.
 CC receptor protein, and fragments and analogues thereof may be
 CC immobilised on a support and used in assay and purification of the
 CC ald target. Labelled Abs, raised to the protein may be injected
 CC into the vascular system to detect the presence of atherosclerotic
 CC plaques.
 CC Sequence 1588 BP; 506 A; 304 C; 367 G; 411 T;

Query Match 4.0%; Score 102.6; DB 1; Length 1588;
 Best Local Similarity 58.8%; Pred. No. 1.4e-15;
 Matches 177; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Oy 680 ttgcgcttctgagagcagcagctgtagcagagccgggtgagccttaccatgctgagc 739
 Db 1058 TCCGACTGTGGTGGTGGACGGCCCTCAGAGGAGAGGAGATTTTTCAGAGAGGCC 1117
 Oy 740 agtgggagacgcttctgtagcagcagcagcagcagcagcagcagcagcagcagcagc 799
 Db 1118 AGTGGGATACGCTGTGTGACGACCGCTGGGAACTGCTGAGAGACAGTGTCTCAGAGA 1177
 Oy 800 agctggcctcagtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagc 859

Db 1178 GCTTGGATACAAAGGTGTTCAAAGTGTGATACAGCACTTATTGGAAGAGTACGG 1237
 Oy 860 gcccaattatgtttagaagtagcagcagcagcagcagcagcagcagcagcagcagcagc 919
 Db 1238 GTCCATATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1297
 Oy 920 caaagagccttcggggagagcagcagcagcagcagcagcagcagcagcagcagcagc 979
 Db 1298 GAATTGAGACAGTGGGCTGTGAGAGCCTGTTCCAGCAGAGATGCTGGGCTCACTTCA 1357
 Oy 980 c 980
 Db 1358 c 1358

RESULT 7
 ID 028540
 AC 028540
 DT 16-FEB-1993 (first entry)
 DE DNA for bovine sol. scavenger receptor.
 KW Macrophage; endotoxaemia; radiolabelled; toxic; degenerate; ss.
 OS Bos taurus.
 FH Key
 FT cds
 Location/Qualifiers
 cds 4.1362
 /*tag= a

FN WO9214482-A
 PD 03-SEP-1992.
 PF 21-FEB-1992; U01370.
 PR 22-FEB-1991; US-662227.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Krieger M;
 DR WPI: 92-315935/38.
 PT Treatment and diagnosis of endotoxaemia and related disease
 PT states - using a polypeptide fragment of the extracellular
 PT portion of a macrophage scavenger receptor protein
 PS Disclosure: Page 36; 56pp; English.
 CC The bovine scavenger receptor protein gene was isolated by screening
 CC a bovine lung cDNA library with degenerate probes designed based on
 CC fragments of the purified scavenger receptor protein. The DNA encodes
 CC a scavenger receptor protein which has a binding capacity for
 CC acetylated low density lipoprotein (LDL) and binds endotoxin.
 CC Fragments of the extracellular portion of the scavenger receptor
 CC protein inactivate endotoxin-related substances and are used in the
 CC diagnosis and treatment of endotoxaemia. They may be administered
 CC to patients at high risk of symptomatic or endotoxic shock. The
 CC fragments may also be fixed to inert supports for purification purposes.
 CC The fragments may be modified to have greater binding affinity for
 CC the endotoxin-related substance than the native scavenger receptor
 CC protein, or to more effectively neutralise the toxic or pathogenic
 CC effects of mols. that bind the scavenger receptor protein or of
 CC organisms which express such mols. See also 028536-9.
 CC Sequence 1588 BP; 506 A; 304 C; 367 G; 411 T;

Query Match 4.0%; Score 102.6; DB 1; Length 1588;
 Best Local Similarity 58.8%; Pred. No. 1.4e-15;
 Matches 177; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Oy 680 ttgcgcttctgagagcagcagctgtagcagagccgggtgagccttaccatgctgagc 739
 Db 1058 TCCGACTGTGGTGGTGGACGGCCCTCAGAGGAGAGGAGATTTTTCAGAGAGGCC 1117
 Oy 740 agtgggagacgcttctgtagcagcagcagcagcagcagcagcagcagcagcagcagc 799
 Db 1118 AGTGGGATACGCTGTGTGACGACCGCTGGGAACTGCTGAGAGACAGTGTCTCAGAGA 1177
 Oy 800 agctggcctcagtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagc 859
 Db 1178 GCTTGGATACAAAGGTGTTCAAAGTGTGATACAGCACTTATTGGAAGAGTACGG 1237
 Oy 860 gcccaattatgtttagaagtagcagcagcagcagcagcagcagcagcagcagcagcagc 919

Db 1238 GTCCATATGGCTGATGAGTATTTTGTTCCTCCCAAGAGTCATCCATTGAAGAGTCA 1297
QY 920 caaagagctcctctgggagagcataactgtgccaataaagaagatctgagtgctcctgta 979
Db 1298 GAATTAGACAGTGGGTGTGTGAGAGCCTGTTGCGACAGAGAAAGATGCTGGGGTCACCTTGCA 1357
QY 980 c 980
Db 1358 c 1358

Search completed: February 27, 2000, 03:42:40
Job time: 2697 sec


```

GENERAL INFORMATION:
APPLICANT: Li, Xiao
APPLICANT: Snyder, Solomon H
TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
TITLE OF INVENTION: Protein Associated with Taste Buds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Banner & Witcoff, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,350B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Moliffe, Susan A
REGISTRATION/DOCKET NUMBER: 33,568
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4360 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
FEATURE:
NAME/KEY: CDS
LOCATION: 94..3963
US-08-470-350B-1

```

```

Query Match          3.8%; Score 97.8; DB 1; Length 4360;
Best Local Similarity 57.3%; Pred. No. 4.5e-17;
Matches 177; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
1330 gggttcctcgtcagactgatgatgagaaataagaagaagcagtgaggtttt 1389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
958 gggttcctcgtcagactgatgatgagaaataagaagaagcagtgaggtttt 1017
1390 atcaatgagccagggggaataatctgtagatgatgatgagcgttaagatgcgcgtg 1449
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1018 taccaggggttcctgaggggtacgctgtgacgacagctgacacacagagagccaacgtg 1077
1450 atctgcctcagctgagctgacaaaggtccctgcagagcaagaacatggcttaattgga 1509
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1078 gcttcgacgacagctgctgctgctgctgctgctgctgctgctgctgctgctgctgga 1137
1510 gaagaaagaaagccatccatctgtagaataatgtagatgacaggaagaaatgaaggtcctg 1569
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1138 caagcctcgtgacatctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1197
1570 gctgactgtagatcaagcaagatattggaagacacaaatgcgcgcagctgaagatgcaga 1629
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1198 tggagctgctccacccagcagctgctgctgctgctgctgctgctgctgctgctgctgga 1257
1630 gttattgt 1638
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1258 gtcattctgt 1266

```

```

RESULT 3
US-08-477-674-9
Sequence 9, Application US/08477674
Patent No. 5644035
GENERAL INFORMATION:
APPLICANT: Kohls, Kirston E.
APPLICANT: Halenbeck, Robert F.
APPLICANT: Taylor, Eric W.
APPLICANT: Wang, Alice M.
APPLICANT: Casipit, Clayton L.
TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: Cetus Oncology Corporation
STREET: 1400 Fifty-Third Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,674
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,714
FILING DATE:
APPLICATION NUMBER: US/07/961,404
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Kenneth M.
REGISTRATION/DOCKET NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 2595.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 420-3152
TELEFAX: (510) 658-5470
TELEX: N/A
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-674-9
Query Match          3.7%; Score 94.4; DB 1; Length 2285;
Best Local Similarity 56.4%; Pred. No. 2.7e-16;
Matches 176; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
678 cattgccttcctggagggcagcagtgatgcatgagggcgggtggaagcttaccatgctgg 737
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 catcgccctgcccagatggggggccacacacacagggcccgctggagatcttataacagag 307
738 ccagtgaggaaacgcttgtagatgacaaatggatgatatgacagatgagaatgtagcag 797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
308 ccagtgaggacactgctgtagacaaactgtagacactgtagatgtagacagcgtgctgccc 367
798 gcagctgggacctcagtgagcattggccaaagcattgcatcagcagcatattttggggaagg 857
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
368 gccctggccttggaagaaagccacacagcctgctggcagacgctgcttggggcagaagatc 427
858 tggccagatattgtagaagatgagcctgacatggaatggaatgagctttcaattgagcagtg 917
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 agcccccattatctgtagacagagctccagtgacagggagacggcctcactgcccagctg 487

```

UY 918 tccaagaagctctctgggagagcacaactgtgtgcccataaagaagtctgtgagctgtcctg 977
| | | | |
DB 488 CAGATCCCTGGGCTGGCTGAAGAGAGCACTGACGACAGAGAGAGACCTGTGTGTCTG 547
| | | | |
QY 978 taccctctaac 989
| | | | |
DB 548 CACCAATGAAGC 559
| | | | |

RESULT 4

US-08-473-791-9
Sequence 9, Application US/08473791
Patent No. 5736340

GENERAL INFORMATION:

APPLICANT: Kohls, Kirston E.
APPLICANT: Halenbeck, Robert F.
APPLICANT: Taylor, Eric W.
APPLICANT: Wang, Alice M.
APPLICANT: Casipit, Clayton L.
TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cetus Oncology Corporation
STREET: 1400 Fifty-Third Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,791
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,714
FILING DATE:
APPLICATION NUMBER: US/07/961,404
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Kenneth M.
REGISTRATION NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 2595.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 420-3152
TELEFAX: (510) 658-5470
TELEX: N/A

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 2285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-473-791-9

Query Match 3.7%; Score 94.4; DB 2; Length 2285;
Best Local Similarity 56.4%; Pred. No. 2.7e-16;
Matches 176; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 678 cattgccttgcctggagcagcagtgctgcatgaaagccggtgtgagctcaccatgctg 737
| | | | |
DB 248 CATTGGCGCTGGCGATGGGGGCGCCACCAACGAGGCGCGGTGAGATCTTACAGAGG 307
| | | | |
QY 738 ccagttgggaacccgtttgtatgacccaatgagatgctccgcatcgaagaagtgtatcag 797
| | | | |
DB 308 CCAATGGGGGACACTGTGTGTACCAACCTGTGTGGACCTGATGATGCCAGCGTCTGTCGG 367
| | | | |
QY 798 gcaagctggcctcagtgagcatgcaccaagcatgacagcatattttgggaagagtc 857
| | | | |

DB 368 GGGCCCTGGGCTTGGAGAACGCCACCGAGCTCTGGGACAGAGCTTGGGCAAGATC 427
| | | | |
QY 858 tggccagttatgttgatgaagtaagctgacgtcgaatggaatggaatggaatggaatggaatg 917
| | | | |
DB 428 AGGCCCCATCATCTGTGAGAGAGTCCAGTGCACGGAGACGAGGCTTCACTGCGCAGCTG 487
| | | | |
QY 918 tccaagaagctctctgggagagcacaactgtgtgcccataaagaagtctgtgagctgtcctg 977
| | | | |
DB 488 CAGATCCCTGGGCTGGCTGAAGAGAGCACTGACGACAGAGAGAGACCTGTGTGTCTG 547
| | | | |
QY 978 taccctctaac 989
| | | | |
DB 548 CACCAATGAAGC 559
| | | | |

RESULT 5

US-08-316-714-9
Sequence 9, Application US/08316714
Patent No. 5965382

GENERAL INFORMATION:

APPLICANT: Kohls, Kirston E.
APPLICANT: Halenbeck, Robert F.
APPLICANT: Taylor, Eric W.
APPLICANT: Wang, Alice M.
APPLICANT: Casipit, Clayton L.
TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cetus Oncology Corporation
STREET: 1400 Fifty-Third Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,714
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/961,404
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Kenneth M.
REGISTRATION NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 2595.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 420-3152
TELEFAX: (510) 658-5470
TELEX: N/A

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 2285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-316-714-9

Query Match 3.7%; Score 94.4; DB 4; Length 2285;
Best Local Similarity 56.4%; Pred. No. 2.7e-16;
Matches 176; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 678 cattgccttgcctggagcagcagtgctgcatgaaagccggtgtgagctcaccatgctg 737
| | | | |
DB 248 CATTGGCGCTGGCGATGGGGGCGCCACCAACGAGGCGCGGTGAGATCTTACAGAGG 307
| | | | |

STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 47..1402
US-08-453-117-1

Query Match 3.7%; Score 93.8; DB 1; Length 2028;
Best local Similarity 56.2%; Pred. No. 3.7e-16;
Matches 176; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 668 cgttccccatcattcgcttctgtgaagcagcagtgatgaagccgggtgagctct 727
1083 CATTTACGAAGTTGCGACTGTGCTGGGAGCGCCCTCAGAGGGGAGGTGAGATAC 1142
728 accatgctggccagtgaggaaaccgtttgtgatgaccaaaggatgataccgatgcagaag 787
DB 1143 TCCACAGCGGCCGCTGGGGTACAAATTGTGACGATGCTGGGAAGTGCAGTTGGACAGG 1202
QY 788 tgatctgcaggcagctgggcccctcagtcagtcacaaagcattgcatagcatatttg 847
DB 1203 TCGTCTGTAGGAGCTTGGGATACCGAGTTCACCGTGCACAAAGCAGCTCACTTTG 1262
QY 848 gggaaaggctggcccagttatgtgataagtaacgctgcactgggaaatgaagcttcaa 907
DB 1263 GACAAAGTACTGTGCAATATGGCTCATGAACTGTTTGTGGAGAGATCATCTA 1322
QY 908 ttgagcagtgctcaaaagagctcctggggagagcataactgtggccataaagaagatgctg 967
DB 1323 TTGAAGAAATGTAAATTCGCAATGGGGAGCAAGAGCCTGTTCACATCTGAAGATGCTG 1382
QY 968 gaatgctctgtac 980
DB 1383 GAGTCACTTGAC 1395

Search completed: February 27, 2000, 21:33:35
Job time: 177 sec

C	8	329.2	12.8	632	60	A1808920	wf66h04.x
C	9	317	12.4	440	42	A1152507	u92a04.x
C	10	314.2	12.3	513	31	AA313027	EST183845
C	11	314	12.3	483	44	A1325429	ms46a12.x
C	12	297.4	11.6	426	49	A1653707	WP3608.x
C	13	296.4	11.6	448	63	A1970451	w10h10.x
C	14	295.2	11.5	449	49	A1624687	t43f05.x
C	15	271	10.6	274	44	AA373034	EST84994
C	16	233.4	9.1	384	28	AA063841	m18105.x
C	17	224.6	9.1	379	62	A1893679	m198c11.y
C	18	223.2	8.8	365	28	AA073513	m198c11.y
C	19	223.2	8.7	303	43	A1195404	u672h11.x
C	20	206.8	8.1	354	44	A1327025	m198c11.x
C	21	181.2	7.1	347	102	A0343432	RPC111-12
C	22	152	5.9	402	33	AA417058	z13c05.x
C	23	127.8	5.0	181	20	T07472	EST05361.Fe
C	24	125.4	4.9	230	70	AV252179	AV252179
C	25	123	4.8	268	50	AV038271	AV038271
C	26	118.8	4.6	163	37	AA023924	mb92d01.x
C	27	117.2	4.6	423	37	AA063308	ae90d07.s
C	28	116.6	4.6	196	70	AV228999	AV228999
C	29	116.4	4.5	424	74	AV206795	UI-H-B11-
C	30	111.6	4.4	215	70	AV247069	AV247069
C	31	106.6	4.2	466	26	W68576	z34c12.x1
C	32	101.8	4.0	449	74	AA184423	f114e03.y
C	33	99.2	3.9	599	49	A1652526	w61d11.x
C	34	98.8	3.9	705	81	B16308	34707.TP.CI
C	35	98.2	3.8	570	79	GA140105	AJ231769
C	36	96.4	3.7	465	26	W88611	zh7906.x1
C	37	95.4	3.8	540	49	AI047474	DXFZ5860
C	38	95	3.7	411	46	A1449832	mt74d08.x
C	39	93.8	3.7	641	51	A1734941	at05d08.x
C	40	92.8	3.6	474	35	AA581208	nd38d06.x
C	41	92.2	3.6	455	50	A1677867	w34d04.x
C	42	90.6	3.5	382	28	Z81180	SSCF12.POR
C	43	89.6	3.5	525	82	AO676495	HS_5512.A
C	44	88.6	3.5	616	45	A1341167	q489d03.x
C	45	86.2	3.4	468	46	A1415453	mc57d06.x

ALIGNMENTS

RESULT 1
A1968581 684 bp mRNA EST 25-AUG-1999
LOCUS wt90c08.x1 NCI-CGAP GC6 Homo sapiens CDNA clone IMAGE:2514734 3'
DEFINITION similar to TR:008762 008762 PROTEASE, SERINE, 12 NEUROTROPIN, ;
RNA sequence.

ACCESSION
A1968581
A1968581.1 GI:5765399

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
1 (bases 1 to 684)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE
Unpublished (1997)

COMMENT
On Jun 22, 1998 this sequence version replaced gi:3246664.

CONTACT: Robert Strausberg, Ph.D.

TEL: (301) 496-1550

EMAIL: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www.bio.lim.gov/db/ftp/image/image.html

Seq primer: -40UP from g1bco

High quality sequence stop: 410.

Location/Qualifiers

FEATURES

source

1. 684

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2514734"

/issue_lib="NCI-CGAP GC6"

/issue_type="pooled germ cell tumors"

/lab_host="PH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Plasmid DNA from the normalized library

NCI-CGAP GC4 was prepared, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from a pool of 5,000 clones made

from the same library (clones 1257096-1258631,

1469064-1470983, and 147592-1476743). Subtraction by

Bento Soares and M. Fatima Bonaldo.

BASE COUNT 164 a 175 c 141 g 204 t

ORIGIN

Query Match

Best Local Similarity 98.3%; Pred. No. 4.2e-134;

Matches 630; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY	1924	tatcatctctgtgtaccagagaggttggagaaatgtgagattgaacattgatt	1983
DB	684	TACCATTTGGGTACACAGAGAGATTGAGAGAAATTCACATGATGGAT	625
QY	1984	catcgagagatcgaccgcgcgc-gcagtgattatgacatagccctgttagattcaag	2042
DB	624	CATCGGAGATCGAACCGACCGCAGTATTCATATGACCTTGTTAGATTCAAG	565
QY	2043	accagaagcaatgtgcagattcagagcca-tgtttgcagccctgttaccactct	2101
DB	564	ACCAGAAGCAATGTGCCAGATTCCAGACCCATGTTTGCAGCCTGTTACCACTCT	505
QY	2102	ggagagagagccacagaaacagatccactgttaccataacagatgggtgacacag	2161
DB	504	GGAAGAGAGAGCCACAGAAACAGATCCATCTTACATACAGATGGGTGACACG	445
QY	2162	gacgagccttcaagaacactcaacacagcagccatcccttacttccataaagtttt	2221
DB	444	GACGAGCCTTTCAAGACACTCAACACAGACCCATTCCTTACTTCTTAAGGTTT	385
QY	2222	gtgaagaagctataaaggtcgggtttacagaggaagatgtgtgtggaacccatg	2281
DB	384	GTAAGAAGCTTATTAAGGTCGGTTTACAGGAGATGCTTGTGTGAACCTTCATG	325
QY	2282	aacacaaacggtgagcagcttgcagggagacagcggaagaccactcatgttgaagcc	2341
DB	324	AACCAAAACCGTGACAGCTGCCAGGAGACGCGGAGACCACTCATGTGACGCGC	265
QY	2342	ccggaagaagctgtggt	2401
DB	264	CCGGAAGAGACTGGTGCTGTATGGGCTACTCTCCGGGGTATGGCTGTGAAGTCAAG	205
QY	2402	atttcctgtgtttataccaaggttcagccctgttacctgtgtataaagttgacaca	2461
DB	204	ATTCCTGTGTGTATTAACCAAGCTCAGCCTTGTACTTGATTAATAAGTGCACCA	145
QY	2462	aactgttaattcttcataagaaactcaaacagcatttaacaaatgtgaacttgaacc	2521
DB	144	AACGTGATTTCTTCATGAAACTTCAAGACACATTTAAACAAATGAAACTTGAAC	85
QY	2522	cccacattagcactcagcagagatatacaacaagcagcag 2562	
DB	84	CCCACTTTTGCACCTACGACGAGATACACAAATGGCAAG 44	

RESULT 2
 LOCUS AI457628/c
 DEFINITION t53e07.x1 Soares_NSF_F8_9W_OT_PA.P_S1 Homo sapiens cDNA clone IMAGE:2145212 3' similar to TM:008762 008762 NEUROTRPFPSIN
 PRECURLOR : mRNA sequence.
 AI457628
 ACCESSION AI457628.1 GI:4310497
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 614)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgp.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On May 18, 1998 this sequence version replaced gi:3138391.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 This clone is available royalty-free through LINTL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1113 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 384.
 Location/Qualifiers
 1..614
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2145212"
 /clone_lib="Soares_NSF_F8_9W_OT_PA.P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site.1: Not I; Site.2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and cloneids: Soares NBHSF pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 1450332-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HR-9W pool 1:
 758280-760583, 772104-774407 Soares NBHP pool 1:
 304776-306311, 320136-322823, 326280-326663 Soares NBHOT
 pool 1: 723720-726407, 739980-740999 Subtraction by Benicio
 Soares and M. Fatima Bernaldo."
 BASE COUNT 146 a 138 c 124 g 204 t 2 others
 ORIGIN
 Query Match 16.4%; Score 419.8; DB 46; Length 614;
 Best Local Similarity 95.8%; Pred. No. 2.6e-90;
 Matches 430; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 QY 2114 cacagaaaacagcatcacaactgttatcataacaggaatgggtgtacacaggaagcatt 2173
 Db 613 CCCAGAAAACAGCATCCANCTGTTACATACACGATGGGTGACACAGCGAGCCTATT 554
 QY 2114 caaagaacactcaaaaacagccattccctactactccaaagaagtgttgaagaagct 2233
 Db 553 CAAAGACACTCAAAACAGCAGCATTCCTTCCATTCTCNTPAAAGGTTTGTGAAGAACGTT 494
 QY 2234 ataaggctcgggtttacagggagaatgctttgtgtcgtgaacacctcatatcaacaacgcg 2293
 Db 493 ATAAGGTCGGTTCACAGGAGAAATGCTTTGTGCTGGAACCTCATATACCAAAACGG 434
 QY 2294 tggacagcttcgaaggagagacagcgaggaaccaactatgtgtgaacgcccggagagagct 2353

Db	433	TGACAGCTGCCAGGGAGACACGGGAGCACACTCATGTGTGAACGC GCCGAAAAAACT	374
OY	2354	gggtggtgatgtagtgagctcctctggggataaggctgttggaaactaatgattcttcctggcg	2413
Db	373	GggttgtatgtatgggtgacctccctctggggatggctgtgaacTCAAGGATTCCTCCGGTG	314
OY	2414	ttaataccaagaatctcaagcctttgaccttggatataaaaagtgcacaaccaactgtaattct	2473
Db	313	TTTATACCAAATCTCAGCGCTTGTTCCTTGATATAAAGGGTCCOCCAACGTATAATT	254
OY	2474	tcatggaactccaagcagcatttaacaatgtgaaaaacttgaaccoccatattag	2533
Db	253	TCATGGAACCTTCAAGCGCATTTTAACAATGGAATAAATTGAACCCCTTATTAC	194
OY	2534	actcagcagatgacgaacaacagcgcaa	2862
Db	193	ACTCAGCAAGAATGCCAACAAATGGCAAG	105
RESULT	3		
AA460641			
LOCUS	AA460641	397 bp	mRNA
DEFINITION	zfx2d12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:796055.5, similar to SW:TRR2_CANFA P06872 TRYPSTINOEN, ANTONIC PRECURSOR; contains Alu repetitive element; mRNA sequence AA460641	EST	-09-JUN-1997
ACCESSION	AA460641		
VERSION	AA460641.1	GI:2185761	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 397) Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacey,M., Le,N., Lennon,G., Maria,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,K. and Wilson,R. WashU-Merck EST Project 1997 Unpublished (1997)		
JOURNAL	On Sep 12, 1996 this sequence version replaced gi:1394273.		
COMMENT	Contact: Wilton RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. Seq primer: -28ml3 rev2 RT from Amersham High quality sequence stop: 279.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..397		
	/organism="Homo sapiens"		
	/db_xref="GDB:6039851"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:796055"		
	/clone_lib="Soares_total_fetus_Nb2HF8_9w"		
	/dev_stage="-8-9 weeks"		
	/lab_host="DH10B"		
	/note="Vector: pUT73D-Pac (Pharmacia) with a modified polynker; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_3: Not I; Site_4: Eco RI; 2nd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_5: Not I; Site_6: Eco RI; 3rd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_7: Not I; Site_8: Eco RI; 4th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_9: Not I; Site_10: Eco RI; 5th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_11: Not I; Site_12: Eco RI; 6th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_13: Not I; Site_14: Eco RI; 7th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_15: Not I; Site_16: Eco RI; 8th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_17: Not I; Site_18: Eco RI; 9th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_19: Not I; Site_20: Eco RI; 10th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_21: Not I; Site_22: Eco RI; 11th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_23: Not I; Site_24: Eco RI; 12th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_25: Not I; Site_26: Eco RI; 13th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_27: Not I; Site_28: Eco RI; 14th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_29: Not I; Site_30: Eco RI; 15th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_31: Not I; Site_32: Eco RI; 16th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_33: Not I; Site_34: Eco RI; 17th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_35: Not I; Site_36: Eco RI; 18th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_37: Not I; Site_38: Eco RI; 19th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_39: Not I; Site_40: Eco RI; 20th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_41: Not I; Site_42: Eco RI; 21st strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_43: Not I; Site_44: Eco RI; 22nd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_45: Not I; Site_46: Eco RI; 23rd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_47: Not I; Site_48: Eco RI; 24th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_49: Not I; Site_50: Eco RI; 25th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_51: Not I; Site_52: Eco RI; 26th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_53: Not I; Site_54: Eco RI; 27th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_55: Not I; Site_56: Eco RI; 28th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_57: Not I; Site_58: Eco RI; 29th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_59: Not I; Site_60: Eco RI; 30th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_61: Not I; Site_62: Eco RI; 31st strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_63: Not I; Site_64: Eco RI; 32nd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_65: Not I; Site_66: Eco RI; 33rd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_67: Not I; Site_68: Eco RI; 34th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_69: Not I; Site_70: Eco RI; 35th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_71: Not I; Site_72: Eco RI; 36th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_73: Not I; Site_74: Eco RI; 37th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_75: Not I; Site_76: Eco RI; 38th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_77: Not I; Site_78: Eco RI; 39th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_79: Not I; Site_80: Eco RI; 40th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_81: Not I; Site_82: Eco RI; 41st strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_83: Not I; Site_84: Eco RI; 42nd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_85: Not I; Site_86: Eco RI; 43rd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_87: Not I; Site_88: Eco RI; 44th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_89: Not I; Site_90: Eco RI; 45th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_91: Not I; Site_92: Eco RI; 46th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_93: Not I; Site_94: Eco RI; 47th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_95: Not I; Site_96: Eco RI; 48th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_97: Not I; Site_98: Eco RI; 49th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_99: Not I; Site_100: Eco RI; 50th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_101: Not I; Site_102: Eco RI; 51st strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_103: Not I; Site_104: Eco RI; 52nd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_105: Not I; Site_106: Eco RI; 53rd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_107: Not I; Site_108: Eco RI; 54th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_109: Not I; Site_110: Eco RI; 55th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_111: Not I; Site_112: Eco RI; 56th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_113: Not I; Site_114: Eco RI; 57th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_115: Not I; Site_116: Eco RI; 58th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_117: Not I; Site_118: Eco RI; 59th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_119: Not I; Site_120: Eco RI; 60th strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_121: Not I; Site_122: Eco RI; 61st strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_123: Not I; Site_124: Eco RI; 62nd strand CDNA was prepared from mRNA obtained from pooled 8-9 week fetal liver tissue; Site_125: Not I; Site_126: Eco RI; 63rd strand CDNA		

Query Match 15.5%; Score 396; DB 34; Length 397;
 Best Local Similarity 100.0%; Pred. No. 1.1e-84;
 Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2152 ggtgacacagagacgacccattcaagaacactcaacaagcagccattcccttacttcc 2211
 |||||
 Db 2 GGTGACACAGAGAGAGCCATTTCAGAACACTCAACAGCAGCATTCCCTTACTTCT 61

OY 2212 aaaaagtttctgtaagaacgttataaaggctcggtttacaggagaatgcttctgctga 2271
 |||||
 Db 62 AAAAGCTTTGTGAAGAACGTTATAGGGTTCGGTTACAGGAGAAATGCTTGTGCTGA 121

OY 2272 aaccctcagaacacaaagcgttgacagctccagaggagagacagcggagagaccatcag 2331
 |||||
 Db 122 AACCTTCAGAACACAAAGCGGTGGACAGCTGCCAGGAGACAGCGAGAGACACATCATG 181

OY 2332 tctgacagcggcggagagagcgtggtgctgctatggggttgaacccctggggggtatggctgt 2391
 |||||
 Db 182 TGTGAACGGCCCGGAGAGAGCTGGGTGTATAGGGGTACCTCTGGGGGTATGGCTGT 241

OY 392 ggaagtcaggatctctctggtgtttataccaaagctcagccttctgataccttgataaa 2451
 |||||
 Db 242 GGAGTCAAGAGATTCTCTGTTTATACCAAGTCTCAGCCTTTGACTTGGATATAA 301

OY 2452 agtgcacccaactgttaattcttcatggaactcaaacagcatttaacaatggaaa 2511
 |||||
 Db 302 AGTGTACCAAACTGTAACTTCTCATGGAAGCTCAAGCAGCATTTAAACAAATGGAAA 361

OY 2512 acctgaacccccactattagcactcagcagagatg 2547
 |||||
 Db 362 ACTTGAACCCCACTATTAGCAGCATCAGCAGAGATG 397

RESULT 4
 AM188298 699 bp mRNA EST 22-NOV-1999
 LOCUS x1394h04.x1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2664919.3' similar to TR:008762 008762 PROTEASE, SERINE, 12
 NEURORTRYPSPIN, ;, mRNA sequence.

ACCESSION AM188298
 VERSION AM188298
 KEYWORDS EST.
 SOURCE AM188298.1 GI:6462734

ORGANISM human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 699)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

REFERENCE 1
 Unpublished (1997)
 On Apr 30, 1999 this sequence version replaced gi:4727469.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

EMAIL: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400P from Glibco
 High quality sequence stop: 475.

FEATURES
 SOURCE 1..699
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2664919"
 /clone_lib="Soares_NFL.T.GBC.S1"
 /lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung Nhlh19W, testis Nhlh1, and B-cell
 NCI-CCAP GCBI) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo.
 BASE COUNT 202 a 144 c 152 g 201 t
 ORIGIN

Query Match 15.0%; Score 383.6; DB 74; Length 699;
 Best Local Similarity 95.4%; Pred. No. 1.3e-81;
 Matches 395; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 2149 tgggtgacacagagacgacccattcaagaacactcaacaagcagccattcccttactt 2208
 |||||
 Db 699 TGGGTACACACAGAGAGAGCCATTACACACATCAACAGCAGCATCCCTTACTT 640

OY 2209 ccttaaaagtttctgtaagaacgttataaaggctcggtttacaggagaatgcttctgct 2268
 |||||
 Db 639 CTTAAAGGTTTGTGAAGAACGTTATAGGGTTCGTTTACAGCGCAATCTTGTGCT 580

OY 2269 ggaacctcctcagaacacaaagcgttgacagctgcagaggagagacggagagaccactc 2328
 |||||
 Db 579 GTAACTTCCATGACACACAAAGCGCTGCACAGCTCCAGGAGAGAGCGGAGACACTC 520

OY 2329 atgtgtgaacggccggagagagcgtggtgctgctatggggtgacccctgggggtatgct 2388
 |||||
 Db 519 ATGTGTGAACGGCCCGGAGAGAGCTGGGTGTATAGGGGTACCTCTGGGGGTATGTC 460

OY 2389 tctgagtgtaaggatctctctggtgtttataccaaagctcagccttctgataccttgata 2448
 |||||
 Db 459 TGTGAGTGAAGAGATTCTCTGTTTATACCAAGTCTCAGCCTTTGACTTGGATA 400

OY 2449 aaaaagtgacccaactgttaattcttcatggaactcaaacagcagcatttaacaatgg 2508
 |||||
 Db 399 AAAAGTGTACCAAACTGTAACTTCTCATGGAAGCTCAAGCAGCATTTAAACAAATGG 340

OY 2509 aaaaacttgaacccccactattagcactcagcagagatgatacaacaagcggcaag 2562
 |||||
 Db 339 AAAACTTGAACCCCACTATTAGCAGCATCAGCAGATGACACAAATGGCAAG 286

RESULT 5
 AM104566 537 bp mRNA EST 20-OCT-1999
 LOCUS x078h09.x1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2603777.3' similar to TR:008762 008762 PROTEASE, SERINE, 12
 NEURORTRYPSPIN, ;, mRNA sequence.

ACCESSION AM104566
 VERSION AM104566
 KEYWORDS EST.
 SOURCE AM104566.1 GI:6075301

ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 537)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

REFERENCE 1
 Unpublished (1997)
 On Mar 10, 1998 this sequence version replaced gi:2948652.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

EMAIL: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400P from Glibco
 High quality sequence stop: 315.

FEATURES
 SOURCE 1..537
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2603777"

/clone.lib="Soares_NFL_T.GBC.S1"
/lab.host="DH10B"

/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung Nhlh19w, testis NHT, and B-cell NCI-CGAP-GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

140 a 130 c 106 g 161 t

Query Match

13.2%; Score 337; DB 69; Length 537;

Best Local Similarity 93.7%; Pred. No. 1.7e-70;

Matches 385; Conservative 0; Mismatches 20; Indels 6; Gaps 3;

QY 2152 ggtgacacggagagccattcaagaacacacacagcagccttcctactctcct 2211
DB 537 GGGGACACAGGACGACGACATTT--AAGAACACTAAACAGCACATTCCTTACT-CT 483
QY 2212 aaaaagtttctgtaagaacgttataaggtctggtttacaagggaaatgcttgcgtga 2271
DB 482 TAAAGGTTTGAAGAGCTTA-AAGGGTCGGTTTACAGGGAGAAATGCTTGTGTGGA 424
QY 2272 aacctcattgaacacaaacgctgtgacagctgcacagggagacagcgagaccactatg 2331
DB 423 AACCTCATGAAACAAACCGCTGGACAGCTGCCAGGAGACAGCGAGACCATCATG 364
QY 2332 tctgacacggccggagagagctgtgtgtatggtgaggtgagccctgggggtatgctgt 2391
DB 363 tctgacacggccggagagagctgtgtgtatggtgaggtgagccctgggggtatgctgt 304
QY 2392 ggaagtcagaagatctcctgtgttataacaaagtcacagccttgccttgcgtga 2451
DB 303 GGAGTCAAGATTCCTCTGCTGTTTATACCAAGTCTACGCTTGTACTTGATGATAA 244
QY 2452 agtgtacacaaactgaattcttcattggaacttcaaacagcatttaacaaatgga 2511
DB 243 AGGTGACCAAACTGTAATCTTCATGGAACCTTCAAGACGATTTAAACAATGGAAA 184
2512 actttgaaccccatatagcactcagcagagatgacacaaacggcgaag 2562
183 ACTTTGAACCCCATTTAGCAGCTCAGAGATGACAAACAATGGCAAG 133

RESULT 6

LOCUS AA928660 478 bp mRNA EST 07-JUL-1998

DEFINITION om75g08.s1 NCI_CGAP_GC4 Homo sapiens CDNA clone IMAGE:1533054 3'

sequence. similar to TR:008762 008762 NEURORTRYPSPIN PRECURSOR.; mRNA

ACCESSION AA928660

VERSION AA928660.1 GI:3076951

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 478)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

On Jan 19, 1998 this sequence version replaced gi:2045775.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/BLNL at:

www.bio.lnlnl.gov/bbrp/image/image.html

Insert Length: 616 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 382.

Location/Qualifiers

1..478

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="IMAGE:1533054"

/clone.lib="NCI_CGAP_GC4"

/tissue_type="pooled germ cell tumors"

/lab.host="DH10B"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from 3 pooled

germ cell tumors, and was then primed with a Not I

oligo(dt) primer. Double-stranded cDNA was ligated to Eco

RI adaptors (Pharmacia), digested with Not I and cloned

into the Not I and Eco RI sites of the modified p773

vector. Library is normalized. Library was constructed by

Bento Soares and M. Fatima Bonaldo."

BASE COUNT

131 a 123 c 90 g 134 t

ORIGIN

Query Match 13.0%; Score 333.8; DB 40; Length 478;

Best Local Similarity 99.1%; Pred. No. 9.7e-70;

Matches 346; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 2215 aggttttgaagaacgttataaaggctgggtttacagggagaatgcttgcgtgaac 2274
DB 478 AGGTTTGTGAAGAACGTTATTAAGGTCGTTTACAGGAGAAATGCTTGTCTGGAAC 419
QY 2275 ctccatgaacacaaacgctgtgaca-gctgcagaggaacacagggagcaactatgtg 2333
DB 418 CTCATGAACACAAACGTTGTGACAGCTGCCAGGAGACCGGAGACCATCTATGTG 359
QY 2334 tgaacggccggagagagctgtgtgtatggtgtgacctctgggggtatgctgtg 2393
DB 358 TGAACGGCCCGAGAGAGCTGGGTGTATGGGTGACCTCTGGGGTATGGCTGTGG 299
QY 2394 agtcaagatctcctgtgttataacaaagctcagccttgccttgccttgcctga 2453
DB 298 AGTCAAGATTCCTCTGCTGTTTATACCAAGTCTCAGCTTGTGATGCTGTAAGA 239
QY 2454 tctcaccacaaactgaattcttcattggaacttcaaacagcatttaacaaatgga 2513
DB 238 TCTCAACCAAACTGTAATCTTCATGGAACCTTCAAGCGCATTTAAACAATGGAAA 179
QY 2514 ttggaaccccatatagcactcagcagagatgacacaaacggcgaag 2562
DB 178 TTGGAACCCCATTTAGCAGCTCAGAGATGACAAACAATGGCAAG 130

RESULT 7

LOCUS AA166524 470 bp mRNA EST 19-DEC-1996

DEFINITION ms46a12.r1 life Tech mouse embryo 13 5dpc 10666014 Mus musculus

CDNA clone IMAGE:614590 5' similar to SW:UROT_MOUSE P11214 TISSUE

PLASMINOGEN ACTIVATOR; mRNA sequence.

ACCESSION AA166524

VERSION AA166524.1 GI:1744123

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

WIDEOR (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:09:43 2000; MasPar time 18.99 Seconds
560.722 Million cell updates/sec
Tabular output not generated.

Title: >US-09-147-947-6
Description: (1-822) from US09147947A.pep (1 of 2)
Perfect Score: 6328
Sequence: 1 PPTRRPPPLRPFRPPPPALP.....PGVTKVSAFVPMIKSVTKL 822

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 35.408; Variance 155.210; scale 0.228

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	603	9.5	638	2	US-08-681- Sequence 3, Applicatio	5.78e-44
2	597	9.4	356	2	US-08-681- Sequence 1, Applicati	2.12e-43
3	583	9.2	355	2	US-08-811- Sequence 47, Applicati	4.37e-42
4	580	9.2	430	1	US-07-942- Sequence 3, Applicati	8.35e-42
5	583	9.2	437	2	US-08-811- Sequence 51, Applicati	4.37e-42
6	583	9.2	355	1	US-08-811- Sequence 39, Applicati	4.37e-42
7	573	9.1	355	1	US-08-427- Sequence 6, Applicatio	3.79e-41
8	573	9.1	355	4	5233256-1 Patent No. 5233256.	3.79e-41
9	573	9.1	355	1	US-08-427- Sequence 2, Applicatio	3.79e-41
10	573	9.1	355	1	US-08-217- Sequence 1, Applicatio	3.79e-41
11	573	9.1	355	1	US-08-137- Sequence 1, Applicatio	3.79e-41
12	573	9.1	355	1	US-08-217- Sequence 1, Applicatio	3.79e-41
13	573	9.1	355	1	US-08-217- Sequence 1, Applicatio	3.79e-41
14	573	9.1	355	2	US-08-811- Sequence 45, Applicati	3.79e-41
15	573	9.1	356	1	US-08-427- Sequence 8, Applicatio	1.04e-41
16	573	9.1	356	1	US-08-427- Sequence 4, Applicatio	3.79e-41
17	578	9.1	365	1	US-08-720- Sequence 83, Applicati	1.29e-41
18	578	9.1	363	1	US-08-093- Sequence 83, Applicati	1.29e-41
19	578	9.1	393	2	US-08-560- Sequence 44, Applicati	1.29e-41
20	578	9.1	411	1	US-08-153- Sequence 18, Applicati	1.29e-41
21	578	9.1	411	1	US-08-087- Sequence 1, Applicatio	1.29e-41
22	578	9.1	411	1	US-08-286- Sequence 18, Applicati	1.29e-41
23	578	9.1	431	4	5188829-1 Patent No. 5188829.	1.29e-41

ALIGNMENTS

ID	Sequence 1	STANDARD:	PRT:	638 AA.
24	578 9.1 432 2	US-08-560- Sequence 47, Applicati	1.29e-41	
25	573 9.1 437 2	US-08-811- Sequence 49, Applicati	3.79e-41	
26	573 9.1 472 2	US-08-811- Sequence 63, Applicati	3.79e-41	
27	573 9.1 527 2	5185259-8 Patent No. 5185259.	3.79e-41	
28	573 9.1 527 1	US-07-609- Sequence 16, Applicati	3.79e-41	
29	573 9.1 527 3	PCT-US910 Patent No. 5344773.	3.79e-41	
30	573 9.1 562 4	US-08-811- Sequence 2, Applicatio	3.79e-41	
31	573 9.1 562 2	US-08-811- Sequence 43, Applicati	3.79e-41	
32	573 9.1 562 2	5185259-3 Patent No. 5185259.	3.79e-41	
33	573 9.1 562 4	5200340-2 Patent No. 5200340.	3.79e-41	
34	573 9.1 562 4	5244676-5 Patent No. 5244676.	3.79e-41	
35	573 9.1 562 2	US-08-560- Sequence 50, Applicati	3.79e-41	
36	573 9.1 562 2	US-08-883- Sequence 38, Applicati	3.79e-41	
37	572 9.0 306 2	US-08-560- Sequence 45, Applicati	4.70e-41	
38	572 9.0 331 2	US-08-560- Sequence 46, Applicati	4.70e-41	
39	568 9.0 335 2	US-08-811- Sequence 59, Applicati	1.11e-40	
40	570 9.0 383 2	US-08-558- Sequence 6, Applicatio	7.24e-41	
41	572 9.0 389 2	US-08-811- Sequence 65, Applicati	4.70e-41	
42	572 9.0 430 2	5219569-2 Patent No. 5219569.	4.70e-41	
43	570 9.0 437 2	US-08-811- Sequence 57, Applicati	7.24e-41	
44	568 9.0 527 4	5520913-1 Patent No. 5520913.	1.11e-40	
45	565 8.9 411 2	US-08-560- Sequence 48, Applicati	2.13e-40	

RESULT 1 STANDARD: PRT: 638 AA.

US-08-681-151-3

Sequence 3, Application US/08681151

Patent No. 5869637

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Bandman, Olga

APPLICANT: Braxton, Scott Michael

APPLICANT: Goll, Surya

TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/681,151

FILING DATE: Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0074US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

CC COUNTRY: USA
CC ZIP: 22202
CC COMPUTER READABLE FORM:


```
CC Sequence 51, Application US/08811949
CC Patent No. 5840533
CC GENERAL INFORMATION:
CC APPLICANT: NITWA, MINEO
CC APPLICANT: SAITO, YOSHIMASA
CC APPLICANT: SASAKI, HITOSHI
CC APPLICANT: HAYASHI, MASAKO
CC APPLICANT: NOTANI, JOUJI
CC APPLICANT: KOBAYASHI, MASAKAZU
CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
CC NUMBER OF SEQUENCES: 67
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC CITY: ARLINGTON
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/811,949
CC FILING DATE: 05-MAR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OBLON, NORMAN F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 18-966-0
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-413-3000
CC TELEFAX: 703-413-2220
CC - INFORMATION FOR SEQ ID NO: 51:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 437 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 437 AA; 48707 MW; 965246 CN;
SQ
Query Match 9.2%; Score 583; DB 2; Length 437;
Best Local Similarity 37.6%; Pred. No. 4.37e-42;
Matches 114; Conservative 66; Mismatches 101; Indels 22; Gaps 18;
Db 136 LGKHNVCNRPDDDAKPMCHVKNRRLTWECVPS-CSTGGLROYSOPOFRITIGLFIADI 194
531 IGRHN-CRHS-DAGVICYF-GKKAS-GNSNKESLSSVCGRLRLHRRQRIIGKNSLR 586
195 ASHPQQAIFAKHRRSPGERFLCGILISSCWITLSAACHFOERF--PPHLLTVILGRYYR 252
587 GGMPQVSLRLKSSHGDG-RLLCGATLLSSCWLTAAHCK-RYGNSTRSYAVRG-DYH 643
Db 253 -VPGEEEOKEFEVEKYIVKEFDDDTYNDIALQLKSDSSRCAQESSVTVCLPP-AD 310
644 TLVPEEFEEIEGQVIVHREYRPDRSDYDIALVRLQGPBEQCARFSSHVLPACLPIMRE 703
Db 311 LQLPWTTECELSGYGKHEALSPFYSERLKEAHVRLYPSSRCTSOHLNRTVTDNMLCAGD 370
704 RPKQTASNCYITIGWG--DT-GRAYSTRLOQAALPLPKRFC-BERYKGR-FTGRMLCAGN 758
Db 371 TRSGGPOANLHADCGDSGGLVCLNDGAMTLV-GIISWGLCGGOKDVPGYTKVTNYLD 429
759 LHE---HKRV-DSQCGDSGGLPMLCERPGESWVYGVTSWYGGCGVADSPGYTKVSAFVP 814
Db 430 WIR 432
815 WIR 817
RESULT 6
CC Sequence 39, Application US/08811949
CC Patent No. 5840533
CC GENERAL INFORMATION:
CC APPLICANT: NITWA, MINEO
CC APPLICANT: SAITO, YOSHIMASA
CC APPLICANT: SASAKI, HITOSHI
CC APPLICANT: HAYASHI, MASAKO
CC APPLICANT: NOTANI, JOUJI
CC APPLICANT: KOBAYASHI, MASAKAZU
CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
CC NUMBER OF SEQUENCES: 67
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC CITY: ARLINGTON
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/811,949
CC FILING DATE: 05-MAR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OBLON, NORMAN F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 18-966-0
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-413-3000
CC TELEFAX: 703-413-2220
CC - INFORMATION FOR SEQ ID NO: 39:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 527 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 527 AA; 59027 MW; 1397140 CN;
SQ
Query Match 9.2%; Score 583; DB 2; Length 527;
Best Local Similarity 37.6%; Pred. No. 4.37e-42;
Matches 114; Conservative 66; Mismatches 101; Indels 22; Gaps 18;
Db 226 LGKHNVCNRPDDDAKPMCHVKNRRLTWECVPS-CSTGGLROYSOPOFRITIGLFIADI 284
531 IGRHN-CRHS-DAGVICYF-GKKAS-GNSNKESLSSVCGRLRLHRRQRIIGKNSLR 586
285 ASHPQQAIFAKHRRSPGERFLCGILISSCWITLSAACHFOERF--PPHLLTVILGRYYR 342
587 GGMPQVSLRLKSSHGDG-RLLCGATLLSSCWLTAAHCK-RYGNSTRSYAVRG-DYH 643
Db 343 -VPGEEEOKEFEVEKYIVKEFDDDTYNDIALQLKSDSSRCAQESSVTVCLPP-AD 400
644 TLVPEEFEEIEGQVIVHREYRPDRSDYDIALVRLQGPBEQCARFSSHVLPACLPIMRE 703
Db 401 LQLPWTTECELSGYGKHEALSPFYSERLKEAHVRLYPSSRCTSOHLNRTVTDNMLCAGD 460
704 RPKQTASNCYITIGWG--DT-GRAYSTRLOQAALPLPKRFC-BERYKGR-FTGRMLCAGN 758
Db 461 TRSGGPOANLHADCGDSGGLVCLNDGAMTLV-GIISWGLCGGOKDVPGYTKVTNYLD 519
```

OY 759 LHE--HKKV-DSCGSDSGPLMCPGSGSVVYVTSWCGCGGVKSDPGVYTKVSAFVP 814
 DB 520 WIR 522
 OY 815 WIR 817

DB 348 WIR 350
 OY 815 WIR 817

Search completed: Mon Mar 13 10:10:06 2000
 Job time : 23 secs.

RESULT 7
 ID US-08-427-640-6 STANDARD: PRT: 355 AA.
 XX xxxxxx
 AC
 DT
 XX

Sequence 6, Application US/08427640
 Patent No. 5658788

GENERAL INFORMATION:

CC APPLICANT: Berg et al.
 CC TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disc
 CC NUMBER OF SEQUENCES: 28
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Eli Lilly and Company
 CC STREET: Lilly Corporate Center
 CC CITY: Indianapolis
 CC STATE: IN.
 CC COUNTRY: U.S.A.
 CC ZIP: 46285
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
 CC COMPUTER: Macintosh
 CC OPERATING SYSTEM: Macintosh
 CC SOFTWARE: Microsoft Word
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/427,640
 CC FILING DATE:
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 07/689,410
 CC FILING DATE: 22 APRIL 1991
 CC INFORMATION FOR SEQ ID NO: 6:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 355 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 355 AA: 38563 MW: 638774 CN;

9.1% Score 573; DB 1; Length 355;

Best Local Similarity 37.3%; Pred. No. 3,79e-41;
 Matches 113; Conservative 66; Mismatches 102; Indels 22; Gaps 18;

DB 54 LGKHNACRNDGDAKPCWCHVLRRLTWEXCDVPS-CSTGGLROYSQPOFRKGLPADT 112
 OY 531 IGRHN-CRHSF-DAGVICDYF-GKKAS-GNSNKSLSVCGILRLHRQRIIGKNSLR 586
 DB 113 ASHPWQAIARHRSRGERFLGGLISSCWLTSAHCGQERE--PRHLLVTLGRYR 170
 OY 587 GGPWQVSLRLKSSHDG-RLLCGATLLSSCWLTAAHCFK-RYGNSTRSYAVRG-DYH 643
 DB 171 -VVGEEQKFEVEKYIVHKEFPDDTYNDIALQLKSDSRCAQESSVRYVCLPP-AD 228
 OY 644 TLVPEFEFEIEIGVQIVIHREYRPDSYDIALVRLQGPREGQCARSSHVLPACLPURE 703
 DB 229 LQLPDTECLSGYKHEALSPYSERLKEAHRLYPSSRCTSQHLNRTVTDNMLCAGD 288
 OY 704 RPKTASNCITITGWC--DT-GRAYSRITLOOAIPLPKRFK-ERYKGR-FTGRMLCAGN 758
 DB 289 TRSGGQANLHDACGSDSGPLMCLNDGRNTLV-GIISWGLGCGKDPGVYTKVNYLD 347
 OY 759 LHE--HKKV-DSCGSDSGPLMCPGSGSVVYVTSWCGCGGVKSDPGVYTKVSAFVP 814

THIS PAGE BLANK (USPTO)

QY 241 ELYHAGOMGVCDQDDDAEVIICRQLGSLGAKAHQAVFEGSGSPVMDDEVCTGNE 300
 Db 301 LSTEOCPKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVVYRGOMGTVC 360
 QY 301 LSTEOCPKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVVYRGOMGTVC 360
 Db 361 DGTETLNTYVVCQQLGKFKYKQASANHFESTGPIWLDVSCSKETRFLOCSRRQGRH 420
 QY 361 DGTETLNTYVVCQQLGKFKYKQASANHFESTGPIWLDVSCSKETRFLOCSRRQGRH 420
 QY 361 DGTETLNTYVVCQQLGKFKYKQASANHFESTGPIWLDVSCSKETRFLOCSRRQGRH 420
 Db 421 DCSHREDVSIACYPGEGGHLISGFVPLMDGKKEGKRVVFIINGOMGTICDDGWTDKD 480
 QY 421 DCSHREDVSIACYPGEGGHLISGFVPLMDGKKEGKRVVFIINGOMGTICDDGWTDKD 480
 Db 421 DCSHREDVSIACYPGEGGHLISGFVPLMDGKKEGKRVVFIINGOMGTICDDGWTDKD 480
 QY 421 DCSHREDVSIACYPGEGGHLISGFVPLMDGKKEGKRVVFIINGOMGTICDDGWTDKD 480
 Db 481 AAVICRQLGKGPAPARTMAVFEKGKPIHYDVKCTGNERSLADCIKODIGRNCRHS 540
 QY 481 AAVICRQLGKGPAPARTMAVFEKGKPIHYDVKCTGNERSLADCIKODIGRNCRHS 540
 Db 541 DAGVTDYFEGKASGNSKESLSVCGRLRHRQKRIIGKNSLRGWPQVSLRLKSS 600
 QY 541 DAGVTDYFEGKASGNSKESLSVCGRLRHRQKRIIGKNSLRGWPQVSLRLKSS 600
 Db 541 DAGVTDYFEGKASGNSKESLSVCGRLRHRQKRIIGKNSLRGWPQVSLRLKSS 600
 QY 541 DAGVTDYFEGKASGNSKESLSVCGRLRHRQKRIIGKNSLRGWPQVSLRLKSS 600
 Db 601 HGDGRLCGATLLSCWVLTAAHCFKRYGNSRSTSAVAVGDIHTLVPEEPEEIGVOQIV 660
 QY 601 HGDGRLCGATLLSCWVLTAAHCFKRYGNSRSTSAVAVGDIHTLVPEEPEEIGVOQIV 660
 Db 601 HGDGRLCGATLLSCWVLTAAHCFKRYGNSRSTSAVAVGDIHTLVPEEPEEIGVOQIV 660
 QY 601 HGDGRLCGATLLSCWVLTAAHCFKRYGNSRSTSAVAVGDIHTLVPEEPEEIGVOQIV 660
 Db 661 IHREYRPRSDYDIALVRLQGEEOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 QY 661 IHREYRPRSDYDIALVRLQGEEOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 Db 661 IHREYRPRSDYDIALVRLQGEEOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 QY 661 IHREYRPRSDYDIALVRLQGEEOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 Db 721 GRAYSTLQOAAIPLLPKRCEERKGRFTGRMLCAGNLEHHRKVDSCOGDSGGLMCR 780
 QY 721 GRAYSTLQOAAIPLLPKRCEERKGRFTGRMLCAGNLEHHRKVDSCOGDSGGLMCR 780
 Db 721 GRAYSTLQOAAIPLLPKRCEERKGRFTGRMLCAGNLEHHRKVDSCOGDSGGLMCR 780
 QY 721 GRAYSTLQOAAIPLLPKRCEERKGRFTGRMLCAGNLEHHRKVDSCOGDSGGLMCR 780
 Db 781 PGESWVYVGTSMGCGVKSDFGVYTKVSAFPMIKSVTKL 822
 QY 781 PGESWVYVGTSMGCGVKSDFGVYTKVSAFPMIKSVTKL 822
 Db 781 PGESWVYVGTSMGCGVKSDFGVYTKVSAFPMIKSVTKL 822
 QY 781 PGESWVYVGTSMGCGVKSDFGVYTKVSAFPMIKSVTKL 822

RESULT 2
 ID W83361 standard; Protein: 875 AA.
 AC W83361:
 DT 17-FEB-1999 (first entry)
 DE Human neurotrophin; tumour inhibition; neurological disease;
 KM Human; neurotrophin; tumour inhibition; neurological disease;
 KM lung disease; gene therapy; drug development; stroke; brain injury;
 KM neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 KM epilepsy; hypoxia; ischemia; nerve transection; neuroangiogenesis;
 KM emphysema; bronchitis.
 KW Homo sapiens.
 PS 5-NOV-1998.
 PR 24-APR-1997: IB0625.
 PA (SOMD)/SONDEREGGER P.
 PI Sondegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72589.
 PT New human and murine neurotrophin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 20-24; 50pp: English.
 CC The present sequence represents human neurotrophin. Neurotrophin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neuroangiogenesis); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain

CC and/or nervous tissue: (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 875 AA:
 Query Match 100.0%; Score 6328; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 54 PTTREPPPLPRFRPPRALPAORPHALQAGHTPRPPMCCPAGBEPVSTYDGAACLRMA 113
 QY 1 PTTREPPPLPRFRPPRALPAORPHALQAGHTPRPPMCCPAGBEPVSTYDGAACLRMA 60
 Db 114 EVPPFLERSPASMAQLRGORHNFCSRPGAGPWCFCYDAGKVDKWCDCRHSVRLR 173
 QY 61 EVPPFLERSPASMAQLRGORHNFCSRPGAGPWCFCYDAGKVDKWCDCRHSVRLR 120
 Db 174 GKRNEFGEIVYASGVWGTSCSHWDDSDASYICHLQLGKGIAKQTPFSGGLIPY 233
 QY 121 GKRNEFGEIVYASGVWGTSCSHWDDSDASYICHLQLGKGIAKQTPFSGGLIPY 180
 Db 234 WSNVRCGDEENILCEKDIWOGVCPQKMAAVTCSFSGPTPIIRLAGSSVHEGRV 293
 QY 181 WSNVRCGDEENILCEKDIWOGVCPQKMAAVTCSFSGPTPIIRLAGSSVHEGRV 240
 Db 294 ELYHAGOMGVCDQDDDAEVIICRQLGSLGAKAHQAVFEGSGSPVMDDEVCTGNE 353
 QY 241 ELYHAGOMGVCDQDDDAEVIICRQLGSLGAKAHQAVFEGSGSPVMDDEVCTGNE 300
 Db 354 LSTEOCPKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVVYRGOMGTVC 413
 QY 301 LSTEOCPKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVVYRGOMGTVC 360
 Db 414 DGTETLNTYVVCQQLGKFKYKQASANHFESTGPIWLDVSCSKETRFLOCSRRQGRH 473
 QY 361 DGTETLNTYVVCQQLGKFKYKQASANHFESTGPIWLDVSCSKETRFLOCSRRQGRH 420
 Db 474 DCSHREDVSIACYPGEGGHLISGFVPLMDGKKEGKRVVFIINGOMGTICDDGWTDKD 533
 QY 421 DCSHREDVSIACYPGEGGHLISGFVPLMDGKKEGKRVVFIINGOMGTICDDGWTDKD 480
 Db 534 AAVICRQLGKGPAPARTMAVFEKGKPIHYDVKCTGNERSLADCIKODIGRNCRHS 593
 QY 481 AAVICRQLGKGPAPARTMAVFEKGKPIHYDVKCTGNERSLADCIKODIGRNCRHS 540
 Db 594 DAGVTDYFEGKASGNSKESLSVCGRLRHRQKRIIGKNSLRGWPQVSLRLKSS 653
 QY 541 DAGVTDYFEGKASGNSKESLSVCGRLRHRQKRIIGKNSLRGWPQVSLRLKSS 600
 Db 654 HGDGRLCGATLLSCWVLTAAHCFKRYGNSRSTSAVAVGDIHTLVPEEPEEIGVOQIV 713
 QY 601 HGDGRLCGATLLSCWVLTAAHCFKRYGNSRSTSAVAVGDIHTLVPEEPEEIGVOQIV 660
 Db 714 IHREYRPRSDYDIALVRLQGEEOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 773
 QY 661 IHREYRPRSDYDIALVRLQGEEOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 Db 774 GRAYSTLQOAAIPLLPKRCEERKGRFTGRMLCAGNLEHHRKVDSCOGDSGGLMCR 833
 QY 721 GRAYSTLQOAAIPLLPKRCEERKGRFTGRMLCAGNLEHHRKVDSCOGDSGGLMCR 780
 Db 834 PGESWVYVGTSMGCGVKSDFGVYTKVSAFPMIKSVTKL 875
 QY 781 PGESWVYVGTSMGCGVKSDFGVYTKVSAFPMIKSVTKL 822

RESULT 3
 ID W83362 standard; Protein: 761 AA.
 AC W83362:
 DT 17-FEB-1999 (first entry)
 DE Mouse neurotrophin.

KW Mouse: neurotrophin; tumour inhibition; neurological disease;
 KW lung disease; gene therapy; drug development; stroke; brain injury;
 KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 KW epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;
 KW emphysema; bronchitis.
 OS Mus musculus.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998; I80625.
 PR 26-APR-1997; CR-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR MPI; 99-009438/01.
 DR N-PSDB; V72390.
 PT New human and murine neurotrophin - used, e.g. for inhibiting
 tumours, treatment of neurological or lung disease, including by
 gene therapy and in drug development
 Claim 1: Page 29-32; 50pp; English.
 CC The present sequence represents mouse neurotrophin. Neurotrophin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularization); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 761 AA:

Query Match 66.4%; Score 4199; DB 1; Length 761;
 Best Local Similarity 87.0%; Pred. No. 0.00e+00;
 Matches 528; Conservative 51; Mismatches 28; Indels 0; Gaps 0;

Db 155 CDCGCGPALPVLIVLVCNGSGHREVELYHAAGVGTICDDQMDNADAVICROLGLSGIAK 214
 QY 216 CFSHSGPTPIRLLAGSSVHEGVELYHAGVGTICDDQMDADAEVLCROLGLSGIAK 275
 Db 215 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 274
 276 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 335
 Db 275 LAGKSTHEGRLEVEYKKGQVGTICDDQMDNADAVICROLGLSGIAK 334
 336 LAGKSTHEGRLEVEYKKGQVGTICDDQMDNADAVICROLGLSGIAK 395
 Db 335 WLDVSCSGKEVFFICSSRWGRHDCSHREDVGLTCYPSDSGHRSLSPGPILVDGENK 394
 QY 396 WLDVSCSGKEVFFICSSRWGRHDCSHREDVGLTCYPSDSGHRSLSPGPILVDGENK 455
 Db 395 KEGREVEFVNGVGTICDDQMDNADAVICROLGLSGIAK 454
 QY 456 KEGREVEFVNGVGTICDDQMDNADAVICROLGLSGIAK 515
 Db 455 CTGNEKALADCVKODIGRHNCRHSEDAVICYLEKKAASSGNKEMLSGCGRLRHRQ 514
 QY 516 CTGNEKALADCVKODIGRHNCRHSEDAVICYLEKKAASSGNKEMLSGCGRLRHRQ 575
 Db 515 KRITGNNLSLRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRYGNSRSY 574
 QY 576 KRITGNNLSLRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRYGNSRSY 635
 Db 575 AVAVGDIHTLVPEEFEEIGVOOIVIHRYNRPDRSDYDIALVRLQGPEDCAFFSSHVLP 634
 QY 636 AVAVGDIHTLVPEEFEEIGVOOIVIHRYNRPDRSDYDIALVRLQGPEDCAFFSSHVLP 695
 Db 635 ACPLWREPRQKATSNCHITGMWDTGRAVSRLTQAAVPLLPFRFCERYKGLFTGRMLC 694

QY 696 ACPLWREPRQKATSNCHITGMWDTGRAVSRLTQAAVPLLPFRFCERYKGLFTGRMLC 755
 Db 695 AGNLQEDNRVDSGCGSGGLMCKEKPDESVMVYGVYSGVCCVKTDPGYTRVPAFVPM 754
 QY 756 AGNLQEDNRVDSGCGSGGLMCKEKPDESVMVYGVYSGVCCVKTDPGYTRVPAFVPM 815
 Db 755 IKSTSL 761
 QY 816 IKSTSL 822

RESULT 4
 ID W99088 standard; Protein: 761 AA.
 AC W99088;
 DE 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 OS Serine protease; BSSP-3; brain tissue.
 OS Mus sp.
 PN W09905290-A1.
 PF 04-FEB-1999.
 PR 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SUNR) SUNTORX LTD.
 PI Tsutoka N, Yamaguchi N, Yamashiro K;
 DR N-PSDB; X19027.
 PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1: Page 51-54; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transformants may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 761 AA:

Query Match 66.4%; Score 4199; DB 1; Length 761;
 Best Local Similarity 87.0%; Pred. No. 0.00e+00;
 Matches 528; Conservative 51; Mismatches 28; Indels 0; Gaps 0;

Db 155 CDCGCGPALPVLIVLVCNGSGHREVELYHAAGVGTICDDQMDNADAVICROLGLSGIAK 214
 QY 216 CFSHSGPTPIRLLAGSSVHEGVELYHAGVGTICDDQMDADAEVLCROLGLSGIAK 275
 Db 215 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 274
 276 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 335
 QY 275 LAGKSTHEGRLEVEYKKGQVGTICDDQMDNADAVICROLGLSGIAK 334
 336 LAGKSTHEGRLEVEYKKGQVGTICDDQMDNADAVICROLGLSGIAK 395
 Db 335 WLDVSCSGKEVFFICSSRWGRHDCSHREDVGLTCYPSDSGHRSLSPGPILVDGENK 394
 QY 396 WLDVSCSGKEVFFICSSRWGRHDCSHREDVGLTCYPSDSGHRSLSPGPILVDGENK 455
 Db 395 KEGREVEFVNGVGTICDDQMDNADAVICROLGLSGIAK 454
 QY 456 KEGREVEFVNGVGTICDDQMDNADAVICROLGLSGIAK 515
 Db 455 CTGNEKALADCVKODIGRHNCRHSEDAVICYLEKKAASSGNKEMLSGCGRLRHRQ 514
 QY 516 CTGNEKALADCVKODIGRHNCRHSEDAVICYLEKKAASSGNKEMLSGCGRLRHRQ 575
 Db 515 KRITGNNLSLRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRYGNSRSY 574
 QY 576 KRITGNNLSLRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRYGNSRSY 635
 Db 575 AVAVGDIHTLVPEEFEEIGVOOIVIHRYNRPDRSDYDIALVRLQGPEDCAFFSSHVLP 634
 QY 636 AVAVGDIHTLVPEEFEEIGVOOIVIHRYNRPDRSDYDIALVRLQGPEDCAFFSSHVLP 695

Db 84 EKEQKVLIOVSCTGTEDTLACQDEE--VYDCSHDEADAGASCENPESFSYPGVRLA 141
 QY 284 EGGSPVLMDEVRCTGNELIEQCPKSSWGEHNGHREDAGVSC-TPLTID-G-V--IRLA 337
 Db 142 DGGHCKGRVEYKHONQWTVCTGMSLRAKVYVROLGCGAVLTLQKCNKHAYGRKI 201
 QY 338 GGGSHGHEGLVEYRQMGKTVCDGTELTNTYVVCROLG-FKYG-KQASANHEESTGPI 395
 Db 202 WLSQMSGSGREATLDDCPSPGKNTCNHDEDTWEC---ED---P--FDRLVGGDNL 252
 QY 396 WLDVSCSGCKETRFLOCSRWGRHDCSHREDVSIACYPGGGSHRLSLGFPVRLMDGENK 455
 Db 253 CGSRLEVLHKGWGSVCDNNGEKEDQVYCKLGCGKSLSPFDRKCYGPGVGTWLDN 312
 QY 456 KEGRVEVFPNGWGTCIDCGMTDKDAVAICROLGY-KGPARA-RIMAYFGGKGPVHVN 513
 313 VRCSGEOSLQCOHRFMGFHDCTHQEDVAVIC 345
 514 VKCTGNERSLADCTIKODIGRHNCRHSEDAVVIC 546
 RESULT
 ID Y13369 standard; Protein: 347 AA.
 AC Y13369:
 DI 25-JUN-1999 (first entry)
 DE Amino acid sequence of protein PRO229.
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 OS Homo sapiens.
 PN W09914328-A2.
 PD 25-MAR-1999.
 PF 16-SEP-1996; U19330.
 PR 25-NOV-1997; US-066840.
 PR 17-SEP-1997; US-059113.
 PR 17-SEP-1997; US-059115.
 PR 17-SEP-1997; US-059117.
 PR 17-SEP-1997; US-059119.
 PR 17-SEP-1997; US-059121.
 PR 17-SEP-1997; US-059122.
 PR 17-SEP-1997; US-059184.
 PR 18-SEP-1997; US-059263.
 PR 18-SEP-1997; US-059266.
 PR 15-OCT-1997; US-062125.
 PR 17-OCT-1997; US-062285.
 PR 17-OCT-1997; US-062287.
 PR 21-OCT-1997; US-062814.
 PR 24-OCT-1997; US-062816.
 PR 24-OCT-1997; US-063045.
 PR 24-OCT-1997; US-063120.
 PR 24-OCT-1997; US-063121.
 PR 24-OCT-1997; US-063127.
 PR 24-OCT-1997; US-063128.
 PR 27-OCT-1997; US-063329.
 PR 27-OCT-1997; US-063327.
 PR 28-OCT-1997; US-063541.
 PR 28-OCT-1997; US-063542.
 PR 28-OCT-1997; US-063544.
 PR 28-OCT-1997; US-063549.
 PR 28-OCT-1997; US-063550.
 PR 28-OCT-1997; US-063564.
 PR 29-OCT-1997; US-063435.
 PR 29-OCT-1997; US-063704.
 PR 29-OCT-1997; US-063732.
 PR 29-OCT-1997; US-063738.
 PR 29-OCT-1997; US-063734.
 PR 29-OCT-1997; US-064215.
 PR 29-OCT-1997; US-063735.

PR 31-OCT-1997; US-063870.
 PR 31-OCT-1997; US-064103.
 PR 03-NOV-1997; US-064248.
 PR 07-NOV-1997; US-064809.
 PR 12-NOV-1997; US-065186.
 PR 17-NOV-1997; US-065846.
 PR 18-NOV-1997; US-065693.
 PR 21-NOV-1997; US-066120.
 PR 21-NOV-1997; US-066364.
 PR 24-NOV-1997; US-066772.
 PR 24-NOV-1997; US-066466.
 PR 24-NOV-1997; US-066770.
 PR 24-NOV-1997; US-066511.
 PR 24-NOV-1997; US-066453.
 PA (Geth J.) GENENTECH INC.
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 DR WPI: 98-229533/19.
 DR N-PSDB: X52240.
 PT New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 PS Claim 12; Fig 54, 320pp; English.
 CC Y13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to growth
 CC or survival of nerve cells including Parkinson's disease, Alzheimer's
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata. PRO266 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair. PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 SQ Sequence 347 AA:

Query Match 11.5% Score 729; DB 1; Length 347;
 Best Local Similarity 40.5%; Pred No. 3 12e-62;
 Matches 135; Conservative 64; Mismatches 110; Indels 24; Gaps 14;

Db 24 VRLVGLHRCRGEVVEEKGQMGTVCDGMDIKDAVAVLCRELDCGASGTPGILYEPRA 83
 QY 227 IRLAGSSVHEGRVELYHAGQMGTVCDGMDADAEVICRQYG-LSGIKAKAHQAY--FG 283
 Db 84 EKEQKVLIOVSCTGTEDTLACQDEE--VYDCSHDEADAGASCENPESFSYPGVRLA 141
 QY 284 EGGSPVLMDEVRCTGNELIEQCPKSSWGEHNGHREDAGVSC-TPLTID-G-V--IRLA 337
 Db 142 DGGHCKGRVEYKHONQWTVCTGMSLRAKVYVROLGCGAVLTLQKCNKHAYGRKI 201
 QY 338 GGGSHGHEGLVEYRQMGKTVCDGTELTNTYVVCROLG-FKYG-KQASANHEESTGPI 395
 Db 202 WLSQMSGSGREATLDDCPSPGKNTCNHDEDTWEC---ED---P--FDRLVGGDNL 252
 QY 396 WLDVSCSGCKETRFLOCSRWGRHDCSHREDVSIACYPGGGSHRLSLGFPVRLMDGENK 455
 Db 253 CGSRLEVLHKGWGSVCDNNGEKEDQVYCKLGCGKSLSPFDRKCYGPGVGTWLDN 312
 QY 456 KEGRVEVFPNGWGTCIDCGMTDKDAVAICROLGY-KGPARA-RIMAYFGGKGPVHVN 513
 Db 313 VRCSGEOSLQCOHRFMGFHDCTHQEDVAVIC 345
 QY 514 VKCTGNERSLADCTIKODIGRHNCRHSEDAVVIC 546

Thu Mar 16 07:57:15 2000

US-09-147-947-6-01.rag

Page 6

Search completed: Mon Mar 13 10:09:25 2000
Job time : 37 secs.

QY 276 AMHQAIVEGSGSPVMDDEVACTGNETLSEIOCPKSSMGEHNGHKEDAGVSCPTLIDGIVR 335
Db 275 LAGGAKTHESRLVYXXGQMGTYCDDGWTENTYVACRLLGFKYKQSSVNHEDGSRPI 334
QY 336 LAGGKSHESRLVYXXGQMGTYCDDGWTENTYVACRLLGFKYKQSSVNHEDGSRPI 395
Db 335 WLDVSCSGKEVEFICQSRMGRHDCSHREDVLTCTYPPSDGRLSPGPRIYVDENK 394
QY 396 WLDVSCSGKEVEFICQSRMGRHDCSHREDVLTCTYPPSDGRLSPGPRIYVDENK 455
Db 395 KEGREVEFVNGWGTICDDGWTDKHAAYICROLGKYPARARTMAYEGEGKPIHMDNVK 454
QY 456 KEGREVEFVNGWGTICDDGWTDKHAAYICROLGKYPARARTMAYEGEGKPIHMDNVK 515
Db 455 CTGNEKALADCVKODIGRHNCRHSEDAAGVICYLKERASSSGNKEMLSSGGCRLHRRQ 514
QY 516 CTGNEKSLADCIKODIGRHNCRHSEDAAGVICYLKERASSSGNKEMLSSGGCRLHRRQ 575
Db 515 KRIGGNNSLRGAMPQASRLRSAGHGRLLCGATILLSGCWVLTAAHCKRKYNNRSY 574
QY 576 KRIGGNNSLRGAMPQASRLRSAGHGRLLCGATILLSGCWVLTAAHCKRKYNNRSY 635
Db 575 AVRVGDTYHLPPEEFOEIGVQOIVIHRYRPRSDYIALVRLQGGECARLSTHVP 634
QY 636 AVRVGDTYHLPPEEFOEIGVQOIVIHRYRPRSDYIALVRLQGGECARLSTHVP 695
Db 635 ACPLMRERPOKTASNCNITGWGDTGRAYSRTLQQAAPLLPKRCKERYKGLFTGMCL 694
QY 696 ACPLMRERPOKTASNCNITGWGDTGRAYSRTLQQAAPLLPKRCKERYKGLFTGMCL 755
Db 695 AGNLOEDNVDSOGSGGGLMCEKPEDESVMVYGVSMGCGCKDPGYTRAPAPVP 754
QY 756 AGNLOEDNVDSOGSGGGLMCEKPEDESVMVYGVSMGCGCKDPGYTRAPAPVP 815
Db 755 IKSVTSU 761
QY 816 IKSVTKL 822

RESULT 2
ENTRY 114893 #type complete
TITLE scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name purple urchin
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSION 114893
REFERENCE 218253
AUTHORS Pancer, Z.; Rast, J.P.; Davidson, E.H.
JOURNAL Immunogenetics (1999) 49:773-786
TITLE Origins of immunity: transcription factors and homologs of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes.
#accession 114893
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2153 #label PAN
#cross-references EMBL:AF064259; NID:g4165052; PID:g4165053; PID:AD08654.1

GENETICS
#gene SRCR12
SUMMARY #length 2153 #molecular-weight 226556 #checksum 3510

Query Match 21.6%; Score 1370; DB 2; Length 2153;
Best Local Similarity 44.5%; Pred. No. 0.00e+00;
Matches 209; Conservative 99; Mismatches 143; Indels 19; Gaps 13;

Db 433 C-PHNGVSHNCAHGEDAG-VSCA-PSSQESRYLVGLNNREGRAVEIFLNNMGTYCDD 489
QY 85 CRSPDAGAPRCGFYCDARGVDMGVCDCRHSVRLRGKNEFGYVAYASGVWGTCCS 144
Db 490 DMGTDPANVVCNOLGIPSSGASARSAYFGKSGPILLDNVCGSGNENSLFLCSNNGICVH 549

QY 145 HMDDSDASVICHODLGCKIAKOTFFSGLGPIYMSVNRKRDENLLEKIMOGG 204
Db 550 NCGHQEDASVYCTGYTTPGMNI-RIVGGSPSEGEVYVGHNGTYCDDMDINDANY 608
QY 205 VCPQKMAAVTCS-FSHGPTPIIRLAGGSSVHEGVELYHAGMGTYCDDQMDADAEV 263
Db 609 VCRELGSAAITSSASQSGDILLDLRCSTGTESSLLTCPRHGVNHCANSEADG 668
QY 264 ICROGLSGIAKAMHOAYTGEESGPMVLDEVRCTGNETLSEIOCPKSSMGEHNGHKEDAG 323
Db 669 VCASTSGPVGTYRVLRYGQNSROGRLEISINMGVPCDDINDATVVCOLGFS 728
QY 324 VSCPTLIDG-VI---RLAGGKSHESRLVYXXGQMGTYCDDGWTENTYVACRL 378
Db 729 SAVAPSAHFGGSGTILMDVSCAGNENSLMDCGHRGLGVNCAHEDAGVCIAS-D 787
QY 379 YG-KQASNHFEESTPIWMLDDVSCGKETRFLQSRMGRHDCSHREDVSIACYPG 437
Db 788 G---PLN--IRLAGRSGHEGVEISLGDDWGTCDSDSGIEDAHVCRQLGF-GPALSA 841
QY 438 GHRSLGFPVRLMDGENKKEGVEVINGQWGTICDDGWTDKDAAVICROLGKQPA-RA 496
Db 842 VTASFGGSGSLMDNVQCSGDEATIAECSHNGIGIHNGCHQEDAGVVC 891
QY 497 RIMAYFEGGKPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAAGVVC 546

RESULT 3
ENTRY 336077 #type complete
TITLE M130 antigen human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Aug-1996
ACCESSION 138003; 336077
REFERENCE 138003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
JOURNAL Eur. J. Immunol. (1993) 23:2320-2325
TITLE A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
#accession 138003
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1116 #label RES
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 43-147
151-254 #domain scavenger receptor cysteine-rich domain homology
258-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology
470-573 #domain scavenger receptor cysteine-rich domain homology
575-678 #domain scavenger receptor cysteine-rich domain homology
711-814 #domain scavenger receptor cysteine-rich domain homology
816-920 #domain scavenger receptor cysteine-rich domain homology
921-1024 #domain scavenger receptor cysteine-rich domain homology
SUMMARY #length 1116 #molecular-weight 120979 #checksum 5687

Query Match 14.9%; Score 946; DB 2; Length 1116;
Best Local Similarity 39.2%; Pred. No. 6.78e-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db	ENTRY	RESULT	4
46	LRLVDGKRCGPREVYQJEOEMGICVNGNSMDEAVSVCINOG - CPYALKAGVANSAG	138006	#type complete
117	VRLLGKNEFEGTIVEYASGWTGVSSSHKDSASVICHQLOLGKGIKOT - PFSGLG	M130 antigen (extracellular variant) - human	
105	SGRI-WMDHVSRCRNESALMDCKHDGKHSNCTHQDAGVTC--DGSNLEM-RLTRGG	#formal_name Homo sapiens #common_name man	
176	LPIIYV-SNVRRCDEENILLEKDLW-QGGVCPOKMAAAYTCSPSHGPTPIILNLAGS	17-May-1996 #sequence_revision 17-May-1996 #text_change	
161	NMCGRIEIKFQ-GRMGTVCDDNFNIDHASVTCROLE-CGSAVSFGSSNFGSGPIWF	02-Aug-1996	
234	SVHEGRVEL-YHAGQMTVCDDMDADAEVLCIRQLGLGSIKAMW-QAYFEFGSGPYML	138006; 635768	
219	DDLCCNNESSALMNCRKQKQKHCNCHADPAGVTCGKADLSLRVDTGECGELERF	138003	
292	DEVACTNEIUSTEIOCPKSSWGSEHNCGRKEDAGVCTPLTDGVIRLAGGSGHEGLREYV	Law, S.K.; Mickle, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;	
279	QGEHGTICDDGWDSYDAVAKOLGCPPTAATAGRNASKGFGHLMDSVSCOGHEPAW	Williams, A.C.; Mason, D.Y.	
352	RQGGTVCDDGWEELNYYVCRQLGRKYGQASAN-HFEESGPIMLDDVSGKGTETFL	Eur. J. Immunol. (1993) 23:2320-2325	
411	QCSRRQGRHDCSHREDVSIACYPGEGHRLSLGFPVRLMDENKKEGVFEVFIINGGT	A new macrophage differentiation antigen which is a member of	
339	QCKHEHGKHYCNHEDAGVTC-SDG-SD-LEL---RLRGSGSRACGTVEYEIQLRGK	the scavenger receptor superfamily.	
471	ICDDGWTDKDAAYTCRLGTYGKPARRTMAYFEGKGPIHYVNVKCTGNERSLACTIKOD	cross-references MUID:93380506	
452	WGGLTCDHYEAKITC 467	cession 138006	
531	IGRNCRHSDAGVTC 546	##status preliminary; translated from GB/EMBL/DBJ	
151-254	##molecule_type mRNA	##residues 1-1149 ##label RES	
258-361	##cross-references EMBL:Z29971; NID:9312147; PID:9312148	##domain scavenger receptor cysteine-rich domain homology	
365-468	##superfamily scavenger receptor cysteine-rich domain homology	##label SRC1\	
470-573	##domain scavenger receptor cysteine-rich domain homology	##label SRC2\	
608-711	##domain scavenger receptor cysteine-rich domain homology	##label SRC3\	
744-847	##domain scavenger receptor cysteine-rich domain homology	##label SRC4\	
849-953	##domain scavenger receptor cysteine-rich domain homology	##label SRC5\	
954-1057	##domain scavenger receptor cysteine-rich domain homology	##label SRC6\	
	##domain scavenger receptor cysteine-rich domain homology	##label SRC7\	
	##domain scavenger receptor cysteine-rich domain homology	##label SRC8\	
	##domain scavenger receptor cysteine-rich domain homology	##label SRC9	
	##length 1149 #molecular_weight 124328 #checksum 487		

Query Match	14.9%	Score 946	DB 2:	Length 1149
Best Local Similarity	39.2%	Prod No. 6,78e-199		
Matches 171:	Conservative	73:	Mismatches 172:	Indels 20: Gaps 16:
Db	46	LRVLDGENKSGREYKVOEEMGTVCNNGMSBAYVLCIONLG-CPTAIKAPGAWNSAG	104	
Oy	117	VRLGKGRNEPEGEVYVYASGVMTVCSSHWDDSDASVICHQLQLGKGIAXQT-PFSLG	175	
Db	105	SGRI-WMDHYSCGNESALMDCKHDSWGKHSNCTEQDQAGVCS--DGSNLEM-RLTNGG	160	
Oy	176	LIPLYW-SNRCRGDEENILLCEKIDW-OGGVCPQKMAAAYCSFSHGPTFPIIRLAGGS	233	
Db	161	NMCGRIEIKFO-GRNGTQDDNFNIDHASYICROLE-CGSVVSFGSSNGEGSPITWF	218	
Oy	234	SVHGREL-YHAGMGVTCDDQWDADAEVLCRQLGLSGIAKANH-QATYEGSGCPML	291	
Db	219	DDLICNGESALMCKCHGNGKHCNCHADAVICSGKADSLRLVDVTECSGLEVRF	278	
Oy	292	DEVACTNETSIHQCRKSSWGHNGCHKEDAVGSTPLTLDGIVIRLAGKSGHEGLVY	351	
Db	279	QGEVGTICDDGMSYDAAVAACKOLGCPYAVTALIGRVNASKGHHITWDSVSCQGHPEVW	338	
Oy	352	RGQMGVTCDDGWTMLNTYVVCRLQGRYKQASAN-HEESTGPTMLDVSQSGKREFL	410	
Db	339	QCKHEWGRKYVCNHNEDAGVTC-SDS-SD-LEL---RLRGGSRCAGTVEYETIORLKG	391	
Oy	411	QCSRRQGRHDCSHREDVYIACPGEGEHRILSLGPVRLMDENNKBEKREVEYFINGMG	470	
Db	392	VCDRGMLKEADVYVCRLQGGGSAKTSYVYSKIAQTNTWFLSSCNGNETSLMDCKNMQ	451	
Oy	471	ICDDGWTDKDAAYICRLQGYKPARAKTAYEGEKGPIHDVNYKCTGNERSLADCIKOD	530	
Db	452	WGGLTCDHYEAKITC	467	
Oy	531	IGRHNCRHSDEAGVTC	546	
RESULT	5			
ENTRY		138004	#type complete	
TITLE		M130 antigen (cytosolic variant 1) - human		
ORGANISM		#formal_name Homo sapiens #common_name man		
DATE		17-May-1996 #sequence_revision 17-May-1996 #text_change 07-Feb-1997		
ACCESSIONS		138004: S36078		
REFERENCE		138003		
#authors		Lay, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.		
#journal		Eur. J. Immunol. (1993) 23:2320-2325		
#title		A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.		
#cross-references		MDID:93380506		
#accession		138004		
#status		preliminary; translated from GB/EMBL/DBJ		
#molecule-type		mRNA		
##residues		1-1151	#label RES	
##cross-references		EMBL:Z29269; NID:g312143; PIR:g312144		
CLASSIFICATION		#superfamily scavenger		
KEYWORDS		cytosol		
FEATURE				
43-147		#domain scavenger receptor cysteine-rich domain	homology	
151-254		#label SRC1\		
		#domain scavenger receptor cysteine-rich domain	homology	
		#label SRC2\		
258-361		#domain scavenger receptor cysteine-rich domain	homology	
		#label SRC3\		
365-468		#domain scavenger receptor cysteine-rich domain	homology	
		#label SRC4\		
470-573		#domain scavenger receptor cysteine-rich domain	homology	
		#label SRC5\		
575-678		#domain scavenger receptor cysteine-rich domain	homology	
		#label SRC6\		
711-814		#domain scavenger receptor cysteine-rich domain	homology	

816-920 #label SRC7\
#domain scavenger receptor cysteine-rich domain homology
#label SRC8\
921-1024 #domain scavenger receptor cysteine-rich domain homology
#label SRC9
#length 1151 #molecular-weight 124820 #checksum 1481

Query Match 14.9%; Score 946; DB 2; Length 1151;
Best Local Similarity 39.2%; Pred. No. 6,786-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVDGENKSGRYEVKVOEEMGTVCNNGSMENAVSYICNOLG-CPTAIKAPGMANSSAG 104
QY 117 VRLGGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHOLQLGKGIKOT-PFSGLG 175
Db 105 SGRI-WMDHYSCRGNEBALMDCKHDKGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 160
QY 176 LPIYV-SNVRCKRDEENILCEKDIW-QGVCVPQKMAAAVTCFSHGPTPIIRLAGGS 233
QY 161 NMCSGRLEIKFO-GRWGTVCDDNFNIDHASVYICROLE-CGSAYFSFGSSNFGSGPIWF 218
234 SYHEGRVEL-YHAGOMGTVCDDQWDADAEVICHOLGLSIGIAKAWH-QAYFEGSGPYML 291
Db 219 DDLICNGNEBALNCKHQGKHNCHDAEDAGVICSKADSLRLVDGVTSCSGLRYRF 278
QY 292 DEVRCTGENELSTIQCPRSSWGEHNGKHKEDAGVSCPTLDVIRLAGGKSGHEGRLEYV 351
Db 279 OGEMGTICDDGMPDSDAAYVACVCLGCPATAVTAIGRVNASKGFHIMLDSVSCOGHEPAWV 338
QY 352 RGMGTVCDDGWTLENTYVVCROLGFKYKQASAN-HFEESTGPIWLDVSCSGKETREL 410
QY 339 QCKHHEMGKHYCNHNEAGVTC-SDG-SD-LEL---RLRGGSGRCAGTVEVEIORLLGK 391
QY 411 QCSRQWGRHDCSHREDVSIACYPGEGHRLSLGFPVRLMDGENKKEGVEVFIINGOMGT 470
Db 392 VCDRGMLKADVYVCRQLGCGSALKTSYQYYSKIQTATNWLFLSSCNGENELMDCKNQ 451
QY 471 ICDDGWTDKDAAYVCRQLGKGPARTMAFYEGEGKPIHVDNVKCTGNESSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRNCRHSEDAGYIC 546

RESULT 6
ENTRY I38005 #type complete
TITLE M130 antigen (cytosolic variant 2) - human
ORGANISM #formal name Homo sapiens #common name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change
07-Feb-1997

IONS
#authors Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
138003
138005; S36079
#journal Eur. J. Immunol. (1993) 23:2320-2325
#title A new macrophage differentiation antigen which is a member of
the scavenger receptor superfamily.
#cross-references NUID:93380506

#accession I38005
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1156 #label RES
#cross-references EMBL:222970; NID:9312145; PID:9312146
CLASSIFICATION
KEYWORDS
cytosol

FEATURE
43-147 #domain scavenger receptor cysteine-rich domain homology
151-254 #domain scavenger receptor cysteine-rich domain homology
256-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology

470-573 #label SRC4\
#domain scavenger receptor cysteine-rich domain homology
#label SRC5\
575-678 #domain scavenger receptor cysteine-rich domain homology
#label SRC6\
711-814 #domain scavenger receptor cysteine-rich domain homology
#label SRC7\
816-920 #domain scavenger receptor cysteine-rich domain homology
#label SRC8\
921-1024 #domain scavenger receptor cysteine-rich domain homology
#label SRC9

Query Match 14.9%; Score 946; DB 2; Length 1156;
Best Local Similarity 39.2%; Pred. No. 6,786-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVDGENKSGRYEVKVOEEMGTVCNNGSMENAVSYICNOLG-CPTAIKAPGMANSSAG 104
QY 117 VRLGGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHOLQLGKGIKOT-PFSGLG 175
Db 105 SGRI-WMDHYSCRGNEBALMDCKHDKGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 160
QY 176 LPIYV-SNVRCKRDEENILCEKDIW-QGVCVPQKMAAAVTCFSHGPTPIIRLAGGS 233
QY 161 NMCSGRLEIKFO-GRWGTVCDDNFNIDHASVYICROLE-CGSAYFSFGSSNFGSGPIWF 218
234 SYHEGRVEL-YHAGOMGTVCDDQWDADAEVICHOLGLSIGIAKAWH-QAYFEGSGPYML 291
Db 219 DDLICNGNEBALNCKHQGKHNCHDAEDAGVICSKADSLRLVDGVTSCSGLRYRF 278
QY 292 DEVRCTGENELSTIQCPRSSWGEHNGKHKEDAGVSCPTLDVIRLAGGKSGHEGRLEYV 351
Db 279 OGEMGTICDDGMPDSDAAYVACVCLGCPATAVTAIGRVNASKGFHIMLDSVSCOGHEPAWV 338
QY 352 RGMGTVCDDGWTLENTYVVCROLGFKYKQASAN-HFEESTGPIWLDVSCSGKETREL 410
QY 339 QCKHHEMGKHYCNHNEAGVTC-SDG-SD-LEL---RLRGGSGRCAGTVEVEIORLLGK 391
QY 411 QCSRQWGRHDCSHREDVSIACYPGEGHRLSLGFPVRLMDGENKKEGVEVFIINGOMGT 470
Db 392 VCDRGMLKADVYVCRQLGCGSALKTSYQYYSKIQTATNWLFLSSCNGENELMDCKNQ 451
QY 471 ICDDGWTDKDAAYVCRQLGKGPARTMAFYEGEGKPIHVDNVKCTGNESSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRNCRHSEDAGYIC 546

RESULT 7
ENTRY A32751 #type complete
TITLE spectrat receptor precursor - sea urchin (Strongylocentrotus
purpuratus)
ORGANISM #formal name Strongylocentrotus purpuratus #common name
purple urchin
DATE 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change
02-Aug-1996
ACCESSIONS
REFERENCE A32751
A32751
#authors Dangott, L.J.; Jordan, J.E.; Bellet, R.A.; Garbers, D.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:2128-2132
#title Cloning of the mRNA for the protein that crosslinks to the
egg peptide spectrat.
#cross-references NUID:89184581

FEATURE
#accession A32751
#molecule_type mRNA
#residues 1-352/'G',354-532 #label DA2
#cross-references GB:J04518
#note the authors translated the codon CAC for residue 353 as
Gly
#note part of this sequence was confirmed by protein
sequencing

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS membrane protein

FEATURE

1-30 #domain signal sequence #status predicted #label SIG\
31-532 #product sceract receptor #status predicted #label MAT\
40-144 #domain scavenger receptor cysteine-rich domain homology
150-257 #label SRC1\
#domain scavenger receptor cysteine-rich domain homology
261-366 #label SRC2\
#domain scavenger receptor cysteine-rich domain homology
379-485 #label SRC3\
#domain scavenger receptor cysteine-rich domain homology

SUMMARY

#length 532 #molecular-weight 57820 #checksum 5701

Very Match 13.5%; Score 852; DB 2; Length 532;

Local Similarity 33.5%; Pred. No. 4,36e-175; Mismatches 177; Indels 24; Gaps 23;

Conservative 99; Mismatches 177; Indels 24; Gaps 23;

Db 41 GNIRLIHGRTENEGSVEIYHATRWGVCWMMHMANVTCKQLGPPGARQFYRAYF-G 99
115 GSVRLRGKNEFEYGVASVGWGTVCSSHMDDSDASVICHQLQL-GKKGIKQTPFSG 173
QY 100 -AHYTFWVYKMKCLGNETLEDCHRPYRPMCNQMAAGVCELPKDEPGSL-RMIL 157
QY 174 LGILPIYWS-NVRCRGDEENILLC-EKDIWQGVCCPKMAAAVTCSEFSHGPTPIIRLAG 231
Db 158 GDVPNEGTLTFWDGAMGSVCHTDFTPDGNVACROWGYSRGVKSITKDGHPGFTGPPII 217
QY 232 GSSVHEGRVELYHAGOGVTCDDDDDAEYICROLGLS-GIAKAMHQAYPEGSGSPVM 290
Db 218 LDAVDEGTEAHITECNMPTPYOHACPTYHNDVGVCKPNEGDIRLMDGSGPHGRY 277
QY 291 LDEVRCGTGNLSTEQCPKSSWG-EHNGCHK-E-DAGVSCPTLTDGYIRLAGKSHGRLL 347
Db 278 EIVHDDAMGTICDDGMDADANVYCRQAGYRGAVKASGEKEDFGFTWAPIHTSEVWCTG 337
QY 348 EYVYRGQMGVCDGWTCLNTYVVCROLGFKYGRKQASAHFEE-S-T-GPIWLDVSCSG 404
Db 338 VEDRLIDCILRDGMT-HSCYHVEDASVYC-ATDDDDTIEIEPKHTIRYIVMGOGGGRVE 395
QY 405 KETRFLOCSRRQ-WGRHDCSHREDVSTACYPGEGHRLSLGF-PVRL-MDGENKKKEGRVE 461
Db 396 VSLGNGWRVCDPDWSDHEAKTYCYHAGYKMGASRAAGSAEVSAPFDLEAPFIIDGITCS 455
QY 462 VFINGQWGTICDDGWTCDKDAAYICROLGYK-GPARATMAYFGE--G-KGPIHVDNVKCT 517
456 GVENETISQCOMKVSADMTCA-TGDVGVCCE 485
518 GNER-SIADCIKODIGRNCRHSDEAGVID 547

Search completed: Mon Mar 13 10:08:30 2000
Job time : 55 secs.

THIS PAGE BLANK (USPTO)

 M O S E L L
 (TM)

Release 3.1a John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:04:57 2000; MasPar time 33.12 Seconds
 Tabular output not generated. 741.229 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (1.822) from US09147947A.pep (1 of 2)
 Perfect Score: 6328
 Sequence: 1 PTRPPPLPRFPFRPPRALP.....PGVYTKVSAFVPIKSVTKL 822

Scoring table: PAM 150
 Gap 11

Searched: 82229 segs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 50.140; Variance 71.058; scale 0.706

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
1	6328	100.0	875	1 NETR_HUMAN NEUTROTYPIN PRECURSOR	0.00e+00
2	4199	66.4	761	1 NETR_MOUSE NEUTROTYPIN PRECURSOR	0.00e+00
3	852	13.5	532	1 SPER_STRPU EGG PEPTIDE SPEPACT RE	2.16e-200
4	838	13.2	1436	1 WC11_BOVIN ANTIEN WCL.1.	2.38e-196
5	634	10.0	625	1 FA11_HUMAN COAGULATION FACTOR XI	5.45e-198
6	607	9.6	566	1 FA01_BOVIN TISSUE PLASMINOGEN ACT	2.30e-130
7	609	9.6	638	1 KAL_MOUSE PLASMA KALLIKREIN PREC	6.28e-131
8	601	9.5	415	1 ACRO_PIG ACROSIN PRECURSOR (EC	1.13e-118
9	600	9.5	559	1 UROT_RAT TISSUE PLASMINOGEN ACT	2.16e-128
10	603	9.5	638	1 KAL_RAT PLASMA KALLIKREIN PREC	3.09e-129
11	603	9.5	638	1 KAL_MOUSE PLASMA KALLIKREIN PREC	3.09e-129
12	580	9.2	431	1 ACRO_RABIT ACROSIN PRECURSOR (EC	9.09e-123
13	580	9.2	431	1 UROK_HUMAN UROKINASE-TYPE PLASMIN	9.09e-123
14	583	9.2	431	1 ACRO_RAT ACROSIN PRECURSOR (EC	1.31e-113
15	576	9.2	559	1 UROT_MOUSE TISSUE PLASMINOGEN ACT	3.58e-124
16	576	9.1	436	1 ACRO_MOUSE ACROSIN PRECURSOR (EC	1.21e-121
17	573	9.1	562	1 UROT_MOUSE TISSUE PLASMINOGEN ACT	8.37e-121
18	551	8.7	433	1 UROK_PAPCY UROKINASE-TYPE PLASMIN	1.21e-114
19	543	8.6	394	1 URTG_DESRO SALIVARY PLASMINOGEN A	2.07e-112
20	545	8.6	790	1 PLMN_PIG PLASMINOGEN (EC 3.4.21	5.72e-113
21	539	8.5	421	1 ACRO_HUMAN ACROSIN PRECURSOR (EC	2.70e-111
22	530	8.4	333	1 PLMN_CANFA PLASMINOGEN (EC 3.4.21	8.68e-109
23	529	8.4	442	1 UROK_PIG UROKINASE-TYPE PLASMIN	1.65e-108

RESULT	1	STANDARD:	PRT:	875 AA.	ALIGNMENTS
ID	NETR_HUMAN				
AC	P56730:				
DT	15-DEC-1999 (rel. 39, Created)				
DT	15-DEC-1999 (rel. 39, Last sequence update)				
DT	15-DEC-1999 (rel. 39, Last annotation update)				
DE	NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).				
GN	PRSS12.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Homidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	TISSUE=BRIN:				
RX	MEDLINE: 98201705.				
RA	PROBA K., GSCHWEND T.P., SONDEREGGER P.;				
RT	"Cloning and sequencing of the cDNA encoding human neutrotypin.";				
RL	Biochim. Biophys. Acta 1396:143-147(1998).				
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC				
CC	ACTION MAY SUBSEVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH				
CC	LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.				
CC	- SIMILARITY: CONTAINS 4 SRCH DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb-sdb.ch).				
CC	EMBL: AJ001531; CAA04816.1; -				
DR	PROSITE: PS00134; TRYPIN_HIS: 1.				
DR	PROSITE: PS00135; TRYPIN_SER: 1.				
DR	PROSITE: PS00420; SPERACT_RECEPTOR: 3.				
KW	Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.				
FT	SIGNAL	1	20	POTENTIAL.	
FT	CHAIN	21	875	NEUTROTYPIN.	
FT	DOMAIN	23	92	PROLINE-RICH.	
FT	DOMAIN	93	165	KRINGLE.	
FT	DOMAIN	170	271	SRCH 1.	
FT	DOMAIN	280	381	SRCH 2.	
FT	DOMAIN	387	487	SRCH 3.	

SQ	SEQUENCE	761 AA;	84118 MW;	3FC9C4F35 CRC32;
	Query Match	66.4%;	Score 4199;	DB 1; Length 761;
	Best Local Similarity	87.0%;	Pred.	No. 0.00e+00;
	Matches	528; Conservative	51; Mismatches	28; Indels 0; Gaps 0;
Db	155 CDCGGPALPVRLVGGNSGHEGRVELYASAGMGTCDDOMNADVDYCROLGSGIAK	214		
Oy	216 CFSHGPTPIRLLAGGVSSVHGEVLELYHAGMGTVCDOMPDADEVICRQLGSLGIK	275		
Db	215 AMHAQAFEGSGSPILLDEVCRTGNELSTIQCKRSSMGEINCHKEEDAGVSCVPLTDGVR	274		
Oy	276 AMHAQVFGGSGSPVMIDEVRCCTGNELSTIEDCKRSSMGEINCHKEEDAGVSCVPLTDGVR	335		
Db	275 LAGKSTHEBGLRELVYYKGWGTVCDDGWTEMTNYACRLLLGFRKYKGROSSNHFDGSNPET	334		
Oy	336 LAGKSGSHBGLRELVYYRGWGTVCCDDGMWTELNTYVVCROLGFRKYKGROSSANHFEEESTGP	395		
Db	335 WLDVVCSGSKREVSLFOCSRPMGRDCHSREDVGVLTYCYDSDSGHRLSPGFIRLNGENK	394		
Oy	396 WLDVVCSGSKRETFLOCSRPMGRDCHSREDVSLACYGEGGHRLSLGFPVRLMDGENK	455		
Db	395 KEGREVEVNGOGWTICDDGMTDKAHAVICROLGYKGPARTMAVFGEKGCPIMHDNVK	454		
Oy	456 KEGREVEVINGOGWTICDDGMTDKAAVICROLGYKGPARTMAVFGEKGCPIMHDNVK	515		
Db	455 CTGNEKALADVCKODIGRRNCRHSEDAVICYDLEKKASSSGNKEMLSGGCLRLHRRO	514		
Oy	516 CTGNRSRLADCICKODIGRRNCRHSEDAVICYDFKAKASGNKNESLSVCGCLRLHRRO	575		
Db	515 KRITGNNSLRGAMWQASLRASHAGDRLLCGATLLSSCWVLTNAHFVKRYGNNSRY	574		
Oy	576 KRITGKNSLRGAWQVYSRLKSSHGDRLLCGATLLSSCWVLTNAHFVKRYGNNSRY	635		
Db	575 AVRVDYHTLVEEPEEOIEGOVIYIHRNRPDRSDYDALVRLQPGRCARLSTHVLP	634		
Oy	636 AVRVDYHTLVEEPEEBELIGVOOYIYHRYRDRSDYDALVRLQPGRCARLSTHVLP	695		
Db	635 ACPFLMRERPOKTASNCHITTMGDTGRAYSRTLQAAYBLPKRFCKERYKGLFTGRMLC	694		
Oy	696 ACPFLMRERPOKTASNCHITTMGDTGRAYSRTLQAAYBLPKRFCEERYKGLFTGRMLC	755		
Db	695 AGNLQEDNRVDSQQDSGGPFLMCERPDESWMVYVYTSKGTGGCVKDPGVYTRVPAPFVW	754		
Oy	756 AGNLHEHRKRVSQCQDSGGPFLMCERPEGESWMVYVYTSWYGCGVXKPSGVYTKVSAPFW	815		
	755 IKSVTSL 761			
	IIIII I			
	816 IKSVTKL 822			
RESULT	3			
ID	SPEC_STRPU	STANDARD:		
AC	p16264:	PRT: 532 AA.		
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	EGG PEPTIDE SPEPACT RECEPTOR PRECURSOR.			
OS	Strongylocentrotus purpuratus (Purple sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;			
OC	Euechinozoa; Echinoidea; Echinozoa; Strongylocentrotidae;			
OC	Strongylocentrotus.			
RN	(1)			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 477-489.			
RX	MEDLINE: 89184581.			
RA	DANGOTT L.J., JORDAN J.E., BELLET R.A., GARBERS D.L.;			
RT	"Cloning of the mRNA for the protein that crosslinks to the egg			
RT	peptide speract.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).			
CC	-1- FUNCTION: RECEPTOR FOR THE EGG PEPTIDE SPEPACT.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	This SWISS-PROT entry is copyright It is produced through a collaboration			

[illegible]

FT	DISULFID	92	109	BY SIMILARITY.
FT	DISULFID	111	120	BY SIMILARITY.
FT	DISULFID	128	209	BY SIMILARITY.
FT	DISULFID	149	191	BY SIMILARITY.
FT	DISULFID	180	204	BY SIMILARITY.
FT	DISULFID	219	300	BY SIMILARITY.
FT	DISULFID	240	282	BY SIMILARITY.
FT	DISULFID	271	295	BY SIMILARITY.
FT	DISULFID	303	434	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	346	362	BY SIMILARITY.
FT	DISULFID	354	423	BY SIMILARITY.
FT	DISULFID	448	523	BY SIMILARITY.
FT	DISULFID	480	496	BY SIMILARITY.
FT	CARBOHYD	513	541	BY SIMILARITY.
FT	CARBOHYD	487	487	POTENTIAL.
SEQ	SEQUENCE	566 AA.	63701 MM.	59A1E93B CRC32.

Try Match 9.6%; Score 607; DB 1; Length 566;
 Local Similarity 37.8%; Pred. No. 2,30e-130;
 Phe 116; Conservative 62; Mismatches 107; Indels 22; Gaps 17/;

Db	265	LGKHNHCNPNQDQAPMCHVKKDKRLTWEYDQVCYVCGLRQYKKRQFRKGGLEFADIT	324
Oy	531	IGRNH-CHHSE--DAVIDIDYFGKKAAGSNKSESISV-CGLRLHRRKORLIGGKNSLRG	587
Db	325	SHPMQALFVKNRRSPGGRFLCGILLISSCWLVSNAHCOFERYPHHLKVLGRT--YR-381	
Oy	568	GMPMQVSLRLKSSHDG-RLCGATLLSSCWLVAHAFK-RYG-NSTRSAVAVGDYHT	644
Db	382	LVPGEEQTFEVEKYIHKFEFDDTYNDIALHLKSDSLTCARESAVRTICLP--DAS	439
Oy	645	LVPEEFEEIEIGVOQIVIHREYRPRSDVIDIALVRLOGEEQCAFSSHVLPACLPWRER	704
Db	440	LQLPWTCCELSGYCKHSSSPFESERLKEAHVLYSRSTSHLNRVTNNMLCAGD	499
Oy	705	PKRTA-SMCIYTGWC--DTGRAY-SRLIQQAIPLPPLRRF-EER-YKGRFTGMACGN	758
Db	500	TRSGGDHNLHDACGDSGGLPYCKMDKDHMLV-GIISMGCGCRKVPGVYTKVTNYLD	558
Oy	759	LH---EHKRV-DSCOGDSGGLPKERPEBSWVYGVISWIGCGVKRSPGYITKVSAPV	814
Db	559	WIRDNTR	565
Oy	815	WIKSVTK	821

RESULT 7
 ID KAL_MOUSE STANDARD: PRT: 638 AA.
 P26262;
 01-MAY-1992 (Rel. 22, Created)
 01-MAY-1992 (Rel. 22, Last sequence update)
 01-NOV-1997 (Rel. 35, Last annotation update)
 PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KLR3 OR PK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=BALE/C; TISSUE=LIVER;
 RX MEDLINE: 91090844.
 RA SEIDAN N G., SAWYER N., HAMELIN J., MION P., BEAUBIN G.,
 RA BRACHAPA L., ROCHEMONT J., MEIKAY M., CHESTEN M.;
 RT 'Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species.';
 RL DNA Cell Biol. 9:737-748(1990).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADIKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.

Query Match	Score	DB 1	Length	638
Best Local Similarity	40.8%	Pred. No. 6.28e-131		
Matches 100	Conservative	54	Mismatches 73	Indels 18; Gaps 16;
390 RIVGGTNAStLEEMWVSLQVLY-SQTHL-CGSSITIGRQVYTLAFCDFGIPDPVWRI 447				
577 RIIGGKNSLRGGWVQVSLRLKSSHGGORLLCGATLLSSCWVTLAAHCFR--RYGNSTRS 634				

```

QY 635 YAVRVDYHTLVPEEPEEEIGVOQIVIHREYRPDRSDYDIALVRLQGPEDQCARFSSHVL 694
Db 501 PICLP-SKADINTIYNQWVTGNGYTKEOGETONILQKATIPLVNPECQKKYRDYVINK 559
QY 695 PACLPIMRERPOKTASNCYITGMDT-GRAYSR-LOQAIPILPKRFCEERYKGRFTGR 752
Db 560 OMICAGT-KEGG-TDACKGDSGGPLVCKHSG-RWOLVGITSMGEGCGRKDPGVYTKVSE 616
QY 753 -MLCAGNLHEHKKRRVDSGDSGGLMCEBPESWVYGYTSMGSGGCVKDSPGVYTKVSA 811
Db 617 YMDMI 621
QY 812 FVPWI 816

```

h completed: Mon Mar 13 10:05:34 2000
 lme : 37 secs.

THIS PAGE BLANK (USPTO)

QY	ICRQLGLGSIKAMHQAVFGESSGQPVMLDEVKTCSTNELSDGCPRSSMGEHNCCHKEDAG	323
Db	669 VVCASITSGPVVGTITRLVVGONSNOGKLEISINNQMGTVCDDSDINDATVYCRLOLGF	728
QY	324 VSCPLDTGG-VI---RLAGKSGHEGRLELYRRGOMGTVCDDGWTLENTVYVCRLOLGF	378
Db	729 SAVSAPTAHFPOGSGTITLMDVDSACAGENSMISDGGHGLGVHNCNAHEDAGVYCIS-D	787
QY	379 YG-KASANHFEESTGPIITLDDVSCSGKETRPLQCSRRQMRHDCSHREDVSTICYGGE	437
Db	788 G--PLN--IRLAGRSGMEGRVEISLGGDWGTCVDDSMGIEDAHVYCRLOLGF-GPALSA	841
QY	438 GHRISLGGPVRIMDENKKEGVEVFINGOMGTICDDGWTMKDAVYICRQLGYKGA-RA	496
Db	842 VTAAFGGSGSILMDNVQSGDEATITECSHNGICGHNCQHDAGVYC	891
QY	497 RTMAFGGEGKPIHYDNVKTGNEHSIADCIKODIGRHNCRHSSEDAVYC	546
QY	2 PRELIMINARY: PRT: 2043 AA.	
AC	096943.	
DT	01-MAY-1999 (TREMBLrel. 10, Created)	
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)	
DE	SROR DOMAIN, MEMBRANE FORM 2.	
GN	SRCRM2.	
OS	Gedidia cydonium (Sponge).	
OC	Euxariota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;	
OC	Astrophorida; Geodiidae; Geodia.	
RA	SEQUENCE FROM N.A.	
RA	BLUMBACH B., PANCER Z., DIEHL-SELFPERT B., STEFFEN R., MOENKNER J.,	
RA	MUELLER I., MUELLER W.E.G.;	
RT	"The putative sponge aggregation receptor. Isolation and	
RT	characterization of a molecule composed of scavenger receptor	
RT	cysteine-rich domains and short consensus repeats."	
RL	J. Cell Sci. 111:2635-2644(1998).	
DR	EMBL: Y14953; CAN/5175.1; -.	
DR	HSSP: P10998; IYVC.	
DR	PROSITE: PS00420; SPERACT_RECEPTOR; 8.	
DR	SEQUENCE 2043 AA; 220896 MW; 4CE19401 CRC32;	
QY	Query Match 20.3%; Score 1285; DB 5; Length 2043;	
QY	Best Local Similarity 43.4%; Pred. No. 2,876-299;	
QY	Matches 192; Conservative 86; Mismatches 150; Indels 14; Gaps 13;	
Db	823 PIRLRNNGTVITONEGTYEILHNGWSAVACDYGWTEAVVACHMLGFATVRAVYTRSFH	882
QY	116 SVRL-RGKK--NEFEGTYEYVASGWCWGVCSSHMDDSDASVYICHLQDLGGKGIKQIFPS	172
Db	883 GAVDGDITFLDNVRCAGTEREIFDCYHSSYISRNCDFRTOQAGVACTNYTASEYPI-RLVGG	941
QY	173 GLGLPIYMSVVRGDEENILTEKIDMOGVCPOKMAAAYTSCFSHGPFPIIRLAGG	232
Db	942 SGPHGREGREITYOGWGTVCDDSMGQPDADYVYCRQLGTANASRATVAERGRGELTMD	1001
QY	233 SSVEHGRELELHAQOMGTVCDDQMDDAEYICRQLGLSTGIAKMAHQAYPEEGSGPMLD	292
Db	1002 NVATCTGFENSDDECKRSNMGDHNGCHRPEDAAVOCGELLP-IRLRDGSNLEGRVVEYFN	1060
QY	293 EVRCTGNELSTEGCPKSSMGEGHNGHKREDAGVSTPLITGVYIRLAGKSGHEGRLELYYR	352
Db	1061 KTWGTVCDDFEN-DIGDATVYCRLOLGYPEAVPAEFPARFGSGEGPIMLDDVACVGTETSLF	1119
QY	353 GOMGTVCDDGWTTEL-NTVVYCRLOLGFYKGRQASA-NHFEESTGPIMLDDVYSCSGKETRFL	410
Db	1120 ECSSSGLGQHNHCQHPEDAGVECN-DHTTPS-PVT-DLRLAGGTATATGKRAVEIPFNNTWGT	1176
QY	411 QCSRRQMRHDCSHREDVSTICYGGECHRLISLGFVRLMDGKKEGRVVEVFINGQWGT	470

DB	1177	ICDSDSDRDEAVVVCRRYLQGFSSAETALSNGYFAGADPDPOPTIMLDDVDFCGSETTITSLT	1236
QY	411	ICDDGWTKRDAVAITRQIGYKGPAPARARMAFYEGSK-G-PIHVDNVKCTGNERSLADCIK	528
Db	1237	SALGEHNCAHYEDAGVRC-y19	1257
QY	529	QDIGRHNCRHSDAGVICDFG	550
RESULT	3	PRELIMINARY:	PRT: 1036 AA.
ID	097378:		
AC	097378:		
DT	01-MAY-1999 (TREMBLrel. 10, created)		
DT	01-MAY-1999 (TREMBLrel. 10, last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, last annotation update)		
DE	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN PRECURSOR.		
OS	Strongylocentrotus purpuratus (Purple sea urchin).		
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;		
CC	Echinozoa; Echinoidea; Echinodermata; Echinozoa; Echinoidea;		
CC	Strongylocentrotus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	PANCER Z., RAST J.P., DAVIDSON E.H.;		
RT	"Origins of Immunity: Transcription Factors and Effector Genes of the		
RL	Vertebrate Immune System Expressed in Sea Urchin Coelomocytes."		
DR	EMBL; AF076513; AAD05493.1; -		
DR	HSSP; P56682; ICCV.		
KW	Signal; Receptor.		
FT	SIGNAL	1	20
FT	CHAIN	21	1036
SD	SEQUENCE	1036 AA; 109704 MW; CBB8C31 CRC32;	
	Query Match	19.2%;	Score 1216; DB 5; Length 1036;
	Best Local Similarity	45.3%;	Pred. No. 1.15e-280;
	Matches	199; Conservative	84; Mismatches 134; Indels 22; Gaps 14
Db	615	QC-QGATLVGGSSNEABREVEIQ-NGYWGITICDSDWICITDASVYCRMGLFGOGASGAPGS	672
QY	111	DCRHSVRLRGKRNKEFEETVEV-YASGVWGTVCSHMDSDASVYCHQLQV-GGKGIKQ	168
Db	673	AHF-GCGGPIQLDDVCGTGAEOITFDCAHAPFGVHNCAHYEDAGVFC-IA-SQD---VR	726
QY	169	TPFSGLIPIYWSNVRCRGDEENILLCEKDIWOGGVCPOPMAAAVTCSTSHGTPPIIR	228
Db	727	LVGGSNEAGREVEIQYNGVWGITICDSDWIDIDASVYCRMGLFGOGASGAPGSAGOGTGL	786
QY	229	LAGGSVHGAEVELYHAGWGTVDDQWDADAVEICRQLGSLIAKAMHAYRGESGP	288
Db	787	IQLDDVCGTGAEOITFDCAHAPFGVHNCAHYEDAGVCI-A-SQDV-RIVGGSNEABREVE	844
QY	289	VMLDEVRECTGNELSTIEQPKKSWGHHNCGHEDKEDAGVSTPLTDIVIRIAGKSGSHEGRLE	348
Db	845	IQYNGVWGITICDSDWIGITIDANVYVLMGLFGOGASAPGSANHRGOGSTPQLDDVCGTGEQ	904
QY	349	VYIRQWQETVDDGTELENTIVYVNRQGLGFKQKQASAN-HFEESTGPWLDDVCSGKET	407
Db	905	TIFDCAHPFPVHNCAHYEDAGVCIYS---Q-----D-VLYDGSNAABREVEIQYNGV	955
QY	408	RFLQCSRQKQRHDCSHEDVISIACYRGGEGHRLSLGFPARLMDGKKNKEBRVVFINGQ	467
Db	956	WGITICDDEWIDITDANVYCRMGLFGOGASAPGSAGOGTGPITQLDDVCGTGAEOITFDC	1015
QY	468	WGITICDDEWIDITDANVYCRMGLFGOGASAPGSAGOGTGPITQLDDVCGTGAEOITFDC	1015
Db	1016	HPFGVHNCAHYEDAGVCI 1034	
QY	528	KODIGRHNCRHSDAGVIC 546	
RESULT	4	PRELIMINARY:	PRT: 1116 AA.
ID	007898		
AC	007898:		

01-NOV-1996 (Tremblrel. 01, created)
 01-NOV-1996 (Tremblrel. 01, last sequence update)
 01-NOV-1999 (Tremblrel. 12, last annotation update)
 M130 ANTIGEN PRECURSOR.
 GN CD163.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RA [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93380506.
 RA LAW S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
 RA MASON D.Y.;
 RT "A new macrophage differentiation antigen which is a member of the
 RT scavenger receptor superfamily."
 PUBM Eur. J. Immunol. 23:2320-2325(1993).
 21
 SEQUENCE FROM N.A.
 RA RITTER M., BUECHLER C., LANGMANN T., SCHMITZ G.;
 RT "Genomic organization of the human CD163 gene."
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL: Z22968; CAB80541.1; -
 DR EMBL: Y18388; CAB45233.1; -
 DR EMBL: Y18389; CAB45233.1; JOINED.
 DR EMBL: Y18390; CAB45233.1; JOINED.
 DR EMBL: Y18391; CAB45233.1; JOINED.
 DR EMBL: Y18392; CAB45233.1; JOINED.
 DR EMBL: Y18393; CAB45233.1; JOINED.
 DR EMBL: Y18394; CAB45233.1; JOINED.
 DR EMBL: Y18395; CAB45233.1; JOINED.
 DR EMBL: Y18396; CAB45233.1; JOINED.
 DR EMBL: Y18397; CAB45233.1; JOINED.
 DR EMBL: Y18398; CAB45233.1; JOINED.
 DR EMBL: Y18399; CAB45233.1; JOINED.
 DR EMBL: Y18400; CAB45233.1; JOINED.
 DR EMBL: Y18401; CAB45233.1; JOINED.
 DR EMBL: Y18402; CAB45233.1; JOINED.
 DR EMBL: Y18403; CAB45233.1; JOINED.
 DR PRAM: PF00530; SRCR: 9
 DR PRINTS: PR00258; SPBRACRCPTR.
 KW Antigen; Signal.
 KM SIGNAL 1 40 POTENTIAL.
 FT CHAIN 41 1116 M130 ANTIGEN.
 FT CHAIN 1116 AA: 120979 MW: F193FBA CRC32;
 SO SEQUENCE
 Query Match 14.9%; Score 946; DB 4; Length 1116;
 Local Similarity 39.2%; Pred. No. 2,40e-208;
 Indels 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;
 46 LRLVDGENKSGRVEVYQVEEAGTVCNNGNMSEAVSYICNOLG-CPTAIKAPGAWNSAG 104
 117 VRLGKGNKEFGTYEYVAVSGWGVCSHMDSDASVYICHLDLGAGKIAKOT-PSGLG 175
 105 SGR-I-WMDHVSCKGNESALMDCKHDGKHSNCTHOADAGVCS--DGSNIEM-RLTRGG 160
 176 LIPIYV-SNVACRGRDEEILLCEKDIW-QGVCPQKMAAVTCSFSHGPFPIIRLAGGS 233
 161 NMGSGRIEIKRQ-GRMGIVCDNINIDHASYICNOL-CGSAVSFSSGSSNFGESGVIWF 218
 234 SVHEGRVEL-THAOGWGVCDQDMDDAEYICRDLISLGIAKMW-QAYFGESGIVML 291
 219 DDLICNGNESALMNCCKHQGWGNKNCDAEDADGAVTCSKADLSLTVDGATSGRLTVRE 278
 292 DEVACTGNEELSDIECPKPSMSWEHNCGKHKEDAGVCTPLTDGVITRLAGKGSHERLEYY 351
 279 QGEGNTICDDGWDSTYDAAVACKOLGCPYATVATIGRVNASKGFHIMLDSVSCOGHEPAVW 338
 352 RGOAGTGVCDDMTEILNTVYVCRDLGFKYKQASAN-HFEESTGPIWDDVSCSGKETRFL 410
 339 QCKHHEKHKHVCNNEDAGVTC-SDG-SD-DEL-----RLRGGSRCAGYEVEYQRLLGK 391
 411 QCSRWQMRHCHCSHREDVSLACYGSGGEGHRLSLDFPVRLLDGENKKKGRVEVFTNGWGT 470

DB	392	VCRBGMGLKADYVOCROLGCGS	AKTSTSYQVYSKIQATFTWFLSSCGNENISLMDCKNMQ	451
QY	471	ICDDGWDKDAVAICRQIGYKGP	ARATMAVFGEGKPIHVDNYKCTGNERSLADCIKOD	530
Db	452	WGGLTCDHYEBAKTC	467	
QY	531	IGRHNCRHSDEAGYIC	546	
RESULT	5	PRELIMINARY;	PRT: 1149 AA.	
ID	Q07901			
AC	Q07901:			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	M130 ANTIGEN, EXTRACELLULAR VARIANT PRECURSOR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 93380506.			
RA	LAW S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,			
RA	MASON D.Y.:			
RT	"A new macrophage differentiation antigen which is a member of the			
RT	scavenger receptor superfamily."			
RL	Eur. J. Immunol. 23:2320-2325(1993).			
DR	EMBL; Z22971; CA80544.1; -			
DR	PFAM; PF00530; SRCR; 9.			
DR	PRINTS; PR00258; SEPRACRCPTR.			
KM	Antigen; Signal.			
FT	SIGNAL	40	POTENTIAL.	
FT	CHAIN	41 1149	M130 ANTIGEN, EXTRACELLULAR VARIANT.	
SO	SEQUENCE	1149 AA; 124328 MW; 4901C708 CRC32;		
Query Match		14.9%; Score 946; DB 4; Length 1149;		
Best Local Similarity		39.2%; Pred. No. 2,40e-208;		
Matches 171; Conservative		73; Mismatches 172; Indels 20; Gaps 16;		
Db	46	LRVDENKRCGAREVKKQEE	NGTVCCNNGMSEAVSYCNOLG-CPTAIKAPGANSAG	104
QY	117	VRLRGKRKEFGGYEYVAASGWYVCS	SHWDDSAVSYCHOLDLGKGIKQT-PRSGLG	175
Db	105	SGRI-WMDHVSRCRNESALMDCKHDG	GKHSNCTHOADAGYCS--DGSNLEM-RLTRGG	160
QY	176	LPIYW-SNVRCRGDEEILLCEKDIW	-QGVCPOKMAAVTCSFSHGPEPIRLAGGS	233
Db	161	NMCSRIIRIKQ-GRMGVCDN	FNIDHASYTCQLE-CGSNVSFSSSNFGBSGYTW	218
QY	234	SVHERVEL-YHAQMGVTCDDQDDA	DAEYICHLQSLGIAKMW-QAYFGEBSGVM	291
Db	219	DDLJCNESALMNC	KHOGMKHNCDAEDAGVYCSGADLSLRIVGVTEGSRVRF	278
QY	292	DEVRCITGELSIEDCPSSMGEHNC	GKEDAGVSCYPLTGVIRLAGKSSHECRLEVIY	351
Db	279	QEGMGITICDDGWSYDAVA	AKOLGCTATATAGRVNASKGFHIMDSVSCGHEPAW	338
QY	352	RGMQVTCDDGTELTENNYV	CRQIGFYKGAQASNN-HFESTGPIWLDVSCSGKETRL	410
Db	339	QCKHHEGKATCNHEDAGYTC	-SDG-SD-LEL---RLRGGSRCAGIVEIQRLLGK	391
QY	411	QCSRWOMGRHDCSHREPVSI	ACYPGEGHRLSLGFPYRLMDGKKEKREVEFINQMGIT	470
Db	392	VCDRGWGLKADYVOCROLGSG	SAKTSYQVYSKIQATNTWFLSSCGNNTSLMDCKNMQ	451
QY	471	ICDDGWDKDAVAICRQIGYKGP	ARATMAVFGEGKPIHVDNYKCTGNERSLADCIKOD	530
Db	452	WGGLTCDHYEBAKTC	467	
QY	531	IGRHNCRHSDEAGYIC	546	
RESULT	6			

```

ID 007899 PRELIMINARY: PRT: 1151 AA.
AC Q07899;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA LAW S.A., MICKLEW K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: Z22969; CAA80542.1; -.
DR PFM: PF00530; SRCR: 9.
DR PRINTS: PRO0258; SPERACTRCPTR.
DR Antigen: signal.
FI CHAIN 1 40 POTENTIAL.
SI SEQUENCE 1151 AA; 124820 MW; A72EDD2F CRC32; M130 ANTIGEN, CYTOPLASMIC VARIANT 1.

Query Match 14.9%; Score 946; DB 4; Length 1151;
Best Local Similarity 39.2%; Pred. No. 2,406-208;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVGENKSGRVEYKVOEEMGTVCNNGMSMEAVSICNOLG-CPTAIKAPGMANSSAG 104
117 VRLRGKNEFEETVEYASGVWGTVCSSHMDDSDASYICHOLQLGKGIKOT-PFSGIG 175
105 SGRV-WMDHVSORGNESALMDCKHDGKHSNCTHQDAGVTC--DGSNLEM-RLTRGG 160
176 LRLPIW-SVNRGRGDEENILCEKDIW-QGVCPQKMAAAVTCFSHGPTPIRLAGGS 233
161 NMCSRIEIKFO-GRMGVTCDDNFNIDHASYICROLE-CGSAYSPSGSSNFGSGSPITF 218
224 SVHEGRVEL-YHAGQWGTICDDQWMDADAEVICRQLGSLGIAKAMH-QAYFGSGSPVML 291
219 DDLICNGESALMNCKHOGKHNCDHAEDAGVTCGADLSRLVDGTECGRELYRF 278
292 DEVRTGNELSTEQCRKSSWGEHNCHEKEDAGVCTPLTDGIVRLGGSGHEGRLEYV 351
279 OGEMGTICDDGSDSYDAVACQKOLGCPITAVTAIGRYNASKGFHILWDSYSCQGEHPAW 338
352 RGQWGTICDDGTELTNTYVVCRLGFKYKQASAN-HFEESTGPIWLDVYSCSGKETREL 410
339 QCKHHEMGKHYCNHNEADAGVTC-SDG-SD-LEL---RLRGGSRCAGTVEYETORLLGK 391
411 QCSRQWGRHDCSHEDYSIACYPGEGHRLSLGFPVRLMDGKKEGREGYFINGQWGT 470
392 VCDRGMGLKEADVRCRQLGCGSALKTSYQVYSKIQTNTWFLSSCNGNETSLMDCKNMQ 451
471 ICDDGWTDKDAVIRQLGKPARARTMAYFEGEGKPIHVDNVKCTGNERSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRHNCRSESDAGVTC 546

```

```

RESULT 7
ID 007900 PRELIMINARY: PRT: 1156 AA.
AC Q07900;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 2 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN (1)

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA LAW S.A., MICKLEW K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: Z22970; CAA80543.1; -.
DR PFM: PF00530; SRCR: 9.
DR PRINTS: PRO0258; SPERACTRCPTR.
DR Antigen: signal.
FI CHAIN 1 40 POTENTIAL.
SI SEQUENCE 1156 AA; 125352 MW; 287A07A0 CRC32; M130 ANTIGEN, CYTOPLASMIC VARIANT 2.

Query Match 14.9%; Score 946; DB 4; Length 1156;
Best Local Similarity 39.2%; Pred. No. 2,406-208;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

```

```

Db 46 LRLVGENKSGRVEYKVOEEMGTVCNNGMSMEAVSICNOLG-CPTAIKAPGMANSSAG 104
117 VRLRGKNEFEETVEYASGVWGTVCSSHMDDSDASYICHOLQLGKGIKOT-PFSGIG 175
105 SGRV-WMDHVSORGNESALMDCKHDGKHSNCTHQDAGVTC--DGSNLEM-RLTRGG 160
176 LRLPIW-SVNRGRGDEENILCEKDIW-QGVCPQKMAAAVTCFSHGPTPIRLAGGS 233
161 NMCSRIEIKFO-GRMGVTCDDNFNIDHASYICROLE-CGSAYSPSGSSNFGSGSPITF 218
224 SVHEGRVEL-YHAGQWGTICDDQWMDADAEVICRQLGSLGIAKAMH-QAYFGSGSPVML 291
219 DDLICNGESALMNCKHOGKHNCDHAEDAGVTCGADLSRLVDGTECGRELYRF 278
292 DEVRTGNELSTEQCRKSSWGEHNCHEKEDAGVCTPLTDGIVRLGGSGHEGRLEYV 351
279 OGEMGTICDDGSDSYDAVACQKOLGCPITAVTAIGRYNASKGFHILWDSYSCQGEHPAW 338
352 RGQWGTICDDGTELTNTYVVCRLGFKYKQASAN-HFEESTGPIWLDVYSCSGKETREL 410
339 QCKHHEMGKHYCNHNEADAGVTC-SDG-SD-LEL---RLRGGSRCAGTVEYETORLLGK 391
411 QCSRQWGRHDCSHEDYSIACYPGEGHRLSLGFPVRLMDGKKEGREGYFINGQWGT 470
392 VCDRGMGLKEADVRCRQLGCGSALKTSYQVYSKIQTNTWFLSSCNGNETSLMDCKNMQ 451
471 ICDDGWTDKDAVIRQLGKPARARTMAYFEGEGKPIHVDNVKCTGNERSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRHNCRSESDAGVTC 546

```

Search completed: Mon Mar 13 10:07:18 2000
Job time : 86 secs.

KV emphysema; bronchitis.
OS Homo sapiens.
PN WO9849322-A1.
PD 05-NOV-1998.
PE 24-APR-1998; IB0625.
PR 26-APR-1997; CH-000966.
PA (SOND/) SONDEREGGER P.
PI Sonderegger P.
DR WPI: 99-009438/01.
DR N-PSDB: V72589.
PT New human and murine neurotrophin - used, e.g. for inhibiting tumours, treatment of neurological or lung disease, including by gene therapy and in drug development.
PS Claim 1: Page 20-24; 50pp; English.
CC The present sequence represents human neurotrophin. Neurotrophin proteins and polynucleotides can be used: (i) to inhibit tumours, including metastases, e.g. of brain or retina; (ii) to minimise tissue damage caused by stroke or brain injury (having a protective effect on the penumbra zone); (iii) to treat or prevent neurodegeneration, neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to increase survival of damaged neurons (e.g. in cases of hypoxia, ischaemia, nerve transection) and to stimulate regeneration and/or restoration of synapses; (v) to treat or prevent retinal disorders (e.g. degeneration or neovascularisation); (vi) to prevent apoptosis (or other causes of cell death) in the nervous system; (vii) to regenerate brain and/or nervous tissue; (viii) to treat pain; (ix) to improve brain performance, including learning and memory; (x) to treat or prevent a wide range of psychiatric disorders; and (xi) to treat brain or lung injury associated with protease expression (specifically emphysema or bronchitis).
Sequence 875 AA:

Query Match 100.0%; Score 611; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.69e-54;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 93 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQHNFCRSPDGRWCFTY 152
QY 40 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQHNFCRSPDGRWCFTY 99
DB 153 DARGKVDWGYCDC 165
QY 100 DARGKVDWGYCDC 112

RESULT 3
ID W83362 standard; Protein; 761 AA.
AC W83362.
DT 17-FEB-1999 (first entry)
PE 24-APR-1998.
PR 26-APR-1997; CH-000966.
PA (SOND/) SONDEREGGER P.
PI Sonderegger P.
DR WPI: 99-009438/01.
DR N-PSDB: V72590.
PT New human and murine neurotrophin - used, e.g. for inhibiting tumours, treatment of neurological or lung disease, including by gene therapy and in drug development.
PS Claim 1: Page 29-32; 50pp; English.
CC The present sequence represents mouse neurotrophin. Neurotrophin proteins and polynucleotides can be used: (i) to inhibit tumours, including metastases, e.g. of brain or retina; (ii) to minimise tissue damage caused by stroke or brain injury (having a protective effect on the penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to increase survival of damaged neurons (e.g. in cases of hypoxia, ischaemia, nerve transection) and to stimulate regeneration and/or restoration of synapses; (v) to treat or prevent retinal disorders (e.g. degeneration or neovascularisation); (vi) to prevent apoptosis (or other causes of cell death) in the nervous system; (vii) to regenerate brain and/or nervous tissue; (viii) to treat pain; (ix) to improve brain performance, including learning and memory; (x) to treat or prevent a wide range of psychiatric disorders; and (xi) to treat brain or lung injury associated with protease expression (specifically emphysema or bronchitis).
Sequence 761 AA:

Query Match 83.0%; Score 507; DB 1; Length 761;
Best Local Similarity 78.1%; Pred. No. 4.16e-43;
Matches 57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

DB 85 CGAGESMGATNIGVPCILMDEVPFLERSPPASMAELRGQHNFCRSPDGRWCFTY 144
QY 40 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQHNFCRSPDGRWCFTY 99

DB 145 NAQKVDWGYCDC 157
QY 100 DARGKVDWGYCDC 112

RESULT 4
ID W99088 standard; Protein; 761 AA.
AC W99088.
DT 13-MAY-1999 (first entry)
DE Mouse serine protease BSSP-3.
OS Mus sp.
PN WO9905290-A1.
PD 04-FEB-1999.
PE 24-JUL-1998; J03324.
PR 24-JUL-1997; JP-213969.
PA (SUNR) SUNTORY LTD.
PI Tsuruoka N, Yamaguchi N, Yamashiro K;
DR WPI: 99-142942/12.
DR N-PSDB: X19027.
PT New serine protease expressed in brain tissue - used in screening for potential serine protease inhibitors for drug use.
PS Example 1: Page 51-54; 69pp; Japanese.
CC The present sequence is a serine protease designated BSSP-3, which is isolated from mouse brain tissue. Transforms may be used to produce the enzyme or its partial sequences. Products from the present invention are used for screening for potential peptide or non-peptide serine protease inhibitors or expression regulators for use as drugs.
Sequence 761 AA:

Query Match 83.0%; Score 507; DB 1; Length 761;
Best Local Similarity 78.1%; Pred. No. 4.16e-43;
Matches 57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

DB 85 CGAGESMGATNIGVPCILMDEVPFLERSPPASMAELRGQHNFCRSPDGRWCFTY 144
QY 40 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQHNFCRSPDGRWCFTY 99

DB 145 NAQKVDWGYCDC 157
QY 100 DARGKVDWGYCDC 112

RESULT 5
ID W52815 standard; Protein; 439 AA.
AC W52815.
DT 07-JUL-1998 (first entry)
DE Tissue plasminogen activator variant 2.
KW tPA: fibrin-stimulated; clot; treatment; vascular disease;
OS Synthetic.
OS Homo sapiens.

FT	Key	Location/Qualifiers
FT	Disulfide_bond	6..36
FT	Disulfide_bond	34..43
FT	Disulfide_bond	51..62
FT	Disulfide_bond	56..75
FT	Disulfide_bond	75..84
FT	Disulfide_bond	180..261
FT	Modified_site	97
FT	/note= "N-glycosylation"	
FT	Disulfide_bond	114..156
FT	Disulfide_bond	145..169
FT	Disulfide_bond	177..308
FT	Cleavage_site	190
FT	Disulfide_bond	220..236
FT	Disulfide_bond	228..297
FT	Disulfide_bond	322..397
FT	Disulfide_bond	354..370
FT	Disulfide_bond	387..414
FT	Domain	6..36
FT	/note= "finger domain"	
FT	Domain	51..91
FT	/note= "contains the growth factor domain"	
FT	Domain	93..174
FT	/note= "Kringler-2 domain"	
FT	Domain	177..440
FT	/note= "serine proteases domain"	
PN	US5714145-A.	
PD	03-FEB-1998	
PF	02-SEP-1988; 240856.	
PR	24-JUL-1989; US-383608.	
PR	02-SEP-1988; US-240856.	
PR	03-OCT-1991; US-770510.	
PR	06-JUL-1993; US-088451.	
PR	07-JAN-1994; US-179059.	
PR	14-APR-1995; US-422736.	
PR	29-MAR-1996; US-622891.	
PR	17-OCT-1996; US-733353.	
PA	(GETH) GENENTECH INC.	
PI	Anderson S, Bennett WF, Botstein D, Higgins DL,	
PI	Pacani NF, Zoller MJ;	
DR	WPI; 98-129803/12.	
PT	Treatment of vascular conditions or disease - using tissue	
PT	plasminogen activator variant having amino acid substitutions in	
PT	protease domain to increase fibrin specificity	
PS	Claim 5; Page -: 31pp; English.	
CC	Variant W52814-W52817 containing two or more functional regions were	
CC	created from tissue plasminogen activator (tPA)(W52813). The t-PA	
CC	variants have a higher fibrin-stimulated activity than	
CC	fibrinogen-stimulated activity so they will act preferentially at the	
CC	site of a clot and not systemically. They can be used for treating	
CC	vascular diseases and conditions or to prevent fibrin deposition or	
CC	adhesion formation or reformation.	
CC	Note: This sequence is not given in the specification but was created	
CC	from the wildtype t-PA sequence (W52813) disclosed by the inventors.	
CC	Sequence 439 AA;	
SQ		
D6	Query Match	31.9%; Score 195; DB 1; Length 439;
OY	Best Local Similarity	39.4%; Pred. NO. 5,34e-10;
OY	Matches	26; Conservative 15; Mismatches 22; Indels 3; Gaps 3
D6	105 SLTSGASCLPWNMILIGKYVTNQPAAQLGLGHNYCRNPDGAKPMWCHYIKNR-RL 163	
OY	48 SYTDGACPLMAEVPFLER-SPAPSWAOLRGO-RHNFCRSSPGAGPMCFYGDARGKV 105	
D6	164 TWGYCD 169	
OY	106 DWGYCD 111	

Accession	Protein Name	Location/Qualifiers
DW	Tissue plasminogen activator variant 1.	
KV	tPA: fibrin-stimulated; clot; treatment: vascular disease; fibrin deposition; adhesion formation.	
OS	Synthetic.	
OS	Homo sapiens.	
FT	Key	
FT	Disulfide_bond	7..18
FT	Disulfide_bond	12..31
FT	Disulfide_bond	31..40
FT	Disulfide_bond	48..129
FT	Disulfide_bond	69..111
FT	Disulfide_bond	100..124
FT	Disulfide_bond	136..217
FT	Modified_site	140
FT	/note= "N-glycosylation"	
FT	Disulfide_bond	157..199
FT	Disulfide_bond	188..212
FT	Disulfide_bond	220..351
FT	Cleavage_site	233
FT	Disulfide_bond	263..279
FT	Disulfide_bond	271..340
FT	Disulfide_bond	365..440
FT	Disulfide_bond	397..413
FT	Disulfide_bond	430..457
FT	Domain	7..47
FT	/note= "contains the growth factor domain"	
FT	Domain	48..129
FT	/note= "Kringler-1 domain"	
FT	Domain	136..217
FT	/note= "Kringler-2 domain"	
FT	Domain	220..483
FT	/note= "serine protease domain"	
PN	US5714145-A.	
PD	03-FEB-1998.	
PF	02-SEP-1988; 240856.	
PR	24-JUL-1989; US-383608.	
PR	02-SEP-1988; US-240856.	
PR	03-OCT-1991; US-770510.	
PR	06-JUL-1993; US-088451.	
PR	07-JAN-1994; US-179059.	
PR	14-APR-1995; US-422736.	
PR	29-MAR-1996; US-622891.	
PR	17-OCT-1996; US-733353.	
PR	(GERTH) GENENTECH INC.	
PI	Anderson S, Bennett WF, Botstein D, Higgins DL,	
PI	Paoni NF, Zoller MJ;	
PI	WPI: 98-129803/12.	
PT	Treatment of vascular conditions or disease - using tissue	
PT	plasminogen activator variant having amino acid substitutions in	
PT	protease domain to increase fibrin specificity	
PS	Claim 4; Page 7. 31pp; English.	
CC	Variant5 W52814-W52817 containing two or more functional regions were	
CC	created from tissue plasminogen activator (tPA)(W52813). The t-PA	
CC	variants have a higher fibrin-stimulated activity than	
CC	fibrinogen-stimulated activity so they will act preferentially at the	
CC	site of a clot and not systemically. They can be used for treating	
CC	vascular diseases and conditions or to prevent fibrin deposition or	
CC	adhesion formation or reformation.	
CC	Note: This sequence is not given in the specification but was created	
CC	from the wildtype t-PA sequence (W52813) disclosed by the inventors.	
SQ	Sequence 483 AA;	
Query Match	31.9%; Score 195; DB 1; Length 483;	
Best Local Similarity	39.4%; Pred. No. 5.34e-10;	
Matches	26; Conservative 15; Mismatches 22; Indels 3; Gaps 3.	
Db	149 SLTESGASCLPWNMILIGKVTYAQNPSAALGKHNCRNPDGAKPCHYLKNR-RL 207	
Oy	48 SVTDFGACRLWAEVPEFLER-SPPASMAQLRQG-RHNFCRSPDGARPMCFGDARGKV 105	
Db	208 TWEXCD 213	
Oy	106 DMGYCD 111	

THIS PAGE BLANK (USPTO)


```

CC INFORMATION FOR SEQ ID NO: 61:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 354 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 354 AA; 39339 MW; 636420 CN;
S0






Query Match 29.3%; Score 179; DB 2; Length 354;
Best Local Similarity 37.9%; Pred. No. 2.77e-08;
Matches 25; Conservative 15; Mismatches 23; Indels 3; Gaps 3;

Db 21 STESGASCLPWNMILIGKTYTQNSAQALGIGKRNCRNPNGDAKPMCHYLKNR-RL 79
QY 48 SVTFGGAPCLFMALVEPPFLER-SPPASMAQLRQ-RNHFGRSPDGARPMCFYGDARGKV 105
Db 80 TWEXCD 85
QY 106 DWGYCD 111

RESULT 6 STANDARD; PRT: 355 AA.
XX ID US-08-811-949-59
XX AC xxxxxx
XX DT
XX DE
XX SE Sequence 59, Application US/08811949
DE Patent No. 5840533
CC GENERAL INFORMATION:
CC APPLICANT: NINJA, MINEO
CC APPLICANT: SATTO, YOSHIMASA
CC APPLICANT: SASAKI, HITOSHI
CC APPLICANT: HAYASHI, MASAKO
CC APPLICANT: NOTANI, JOUJI
CC APPLICANT: KOBAYASHI, MASAKAZU
CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
CC NUMBER OF SEQUENCES: 67
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
CC ADDRESSEE: P.C.
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC CITY: ARLINGTON
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC COMPUTER READABLE FORM
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/811,949
CC FILING DATE: 05-MAR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OBLON, NORMAN F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 18-966-0
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-413-3000
CC TELEFAX: 703-413-2220
CC INFORMATION FOR SEQ ID NO: 59:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 355 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 355 AA; 39550 MW; 638607 CN;
S0

```


THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Mon Mar 13 10:22:03 2000;      MasPar time 8.65 Seconds
Tabular output not generated.              398.104 Million cell updates/sec
```

```

Title: >US-09-147-947-6
Description: (40-112) from US09147947A pep (2 of 6)
Perfect Score: 611
Sequence: 1 CPAGPWNVSVIDFGAPLCIRM.....RPFCEYDANGKVDGKDCDC 73

```

Scoring table:

Searched: 142080 seqs, 47172406 residues

```
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

```
Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 36.459; Variance 65.448; scale 0.557

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	507	83.0	761	2	JC5759	brain-specific serine	2.33e-88
2	207	33.9	946	1	A47298	ror-related receptor	3.02e-23
3	185	30.3	615	1	KFH012	coagulation factor XI	7.43e-19
4	179	29.3	559	1	A29941	t-plasminogen activat	1.12e-17
5	179	29.3	562	1	UKH07	t-plasminogen activat	1.12e-17
6	175	28.6	716	1	A40332	macrophage-stimulat	6.75e-17
7	174	28.5	812	1	PJBO	plasmin (EC 3.4.21.7)	1.06e-16
8	173	28.3	716	1	JC5061	macrophage-stimulat	1.06e-16
9	172	28.2	812	1	PLMS	plasmin (EC 3.4.21.7)	2.58e-16
10	171	28.0	603	2	S58941	coagulation factor XI	4.04e-16
11	171	28.0	790	1	PLPG	plasmin (EC 3.4.21.7)	4.04e-16
12	169	27.7	593	3	S45281	coagulation factor XI	9.83e-16
13	167	27.3	560	1	JC4795	plasma hyaluronan-bi	2.39e-15
14	165	27.0	558	2	JC5878	plasma hyaluronan-bi	5.78e-15
15	164	26.8	559	1	A35029	t-plasminogen activat	8.98e-15
16	163	26.7	728	1	JH0579	hepatocyte growth fac	1.40e-14
17	161	26.4	711	1	A47135	macrophage-stimulat	3.36e-14
18	161	26.4	728	1	A53564	hepatocyte growth fac	3.36e-14
19	161	26.4	728	1	A60185	hepatocyte growth fac	3.36e-14
20	161	26.4	810	2	I46260	plasmin (EC 3.4.21.7)	3.36e-14
21	159	26.0	710	1	I51283	hepatocyte growth fac	8.07e-14
22	157	25.7	4548	1	S00567	apoptotain(a) (EC 3.4	1.93e-13
23	156	25.5	169	2	A00522	plasmin (EC 3.4.21.7)	2.99e-13

Use mit Abb. 2

Applicants

45	125	20.5	625	1	TBBO	thrombin (EC 3.4.21.5
44	126	20.6	617	2	UKBAY	thrombin (EC 3.4.21.5
43	129	21.1	433	3	UKBAY	u-plasminogen activat
42	131	21.4	618	2	A35837	thrombin (EC 3.4.21.5
41	132	21.6	622	1	TBBU	thrombin (EC 3.4.21.5
40	133	21.6	291	2	138098	u-plasminogen activat
39	134	21.9	433	1	JN0560	u-plasminogen activat
38	135	22.1	120	2	E61545	Plasmin (EC 3.4.21.7)
37	137	22.4	445	2	A61545	u-plasminogen activat
36	138	22.6	442	1	UKPp	u-plasminogen activat
35	139	22.6	431	1	UKHU	u-plasminogen activat
34	139	22.7	394	2	JS0600	u-plasminogen activat
33	140	22.9	433	1	UKMS	Plasmin (EC 3.4.21.7)
32	140	22.9	89	2	A60140	u-plasminogen activat
31	142	23.2	432	1	S18932	u-plasminogen activat
30	146	23.9	460	2	B61545	Plasmin (EC 3.4.21.7)
29	146	23.9	123	2	C61545	Plasmin (EC 3.4.21.7)
28	147	24.1	810	1	PLHU	Plasmin (EC 3.4.21.7)
27	149	24.4	810	2	B30848	heparocyte growth fac
26	149	24.4	411	2	F51285	heparocyte growth fac
25	150	24.5	655	1	A46688	heparocyte growth fac
24	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
23	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
22	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
21	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
20	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
19	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
18	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
17	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
16	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
15	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
14	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
13	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
12	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
11	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
10	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
9	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
8	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
7	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
6	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
5	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
4	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
3	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
2	154	25.2	1420	2	A32869	apolipoprotein(a) (EC
1	154	25.2	1420	2	A32869	apolipoprotein(a) (EC

ALIGNMENTS

ENTRY	TITLE	ORGANISM	DATE	ACCESSIONS	REFERENCE	#authors	#journal	#title	#cross-references
JC5759	brain-specific serine proteinase (EC 3.4.21.-) - mouse	#formal_name Mus musculus #common_name house mouse	24-Jan-1998	#sequence_revision 13-Mar-1998	#text_change 17-Mar-1999	JC5759	JC5759		
						Yamamura, Y.; Yamashita, K.; Tsuruoka, N.; Nakazato, H.;			
						Tsujiyama, A.; Yamaguchi, N.			
						Biochem. Biophys. Res. Commun. 1997, 239:386-392			
						Molecular cloning of a novel brain-specific serine protease with a kringle-like structure and three scavenger receptor cysteine-rich motifs.			
						MDDB:98008848			

CLASSIFICATION #superfamily1 trypsin homology; scavenger receptor
KEYWORDS cysteine-rich domain homology
FEATURE glycoprotein; hydrolase; serine proteinase

Accession	Protein	Domain	Residues	Length	Weight	Checksum
85-157	kringle-like	status predicted	#label KRY			
163-157	scavenger receptor	cysteine-rich domain	homology			
163-266			#label SRC7			
166-266, 273-372,	domain scavenger receptor	cysteine-rich	status			
386-486		predicted	#label SRC			
513-516	domain fujin binding	status predicted	#label FRB			
517-755	domain tyrosin homology	#label TRY				
93, 521, 569	binding site	carbohydrate (asn)	(covalent)	#status		
	predicted					
562, 612, 711	active-site	His, Asp, Ser	#status predicted			
SUMMARY	#length	761	#molecular-weight	84136	#checksum	5449

[illegible]

QY 100 DARGVDMGYCDD 112

RESULT 2

ENTRY A47299 #type complete

TITLE ror-related receptor RTR - Pacific electric ray

CONTAINS protein-tyrosine kinase (EC 2.7.1.112)

ORGANISM #formal_name Torpedo californica #common_name Pacific electric ray

DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS A47299

REFERENCE A47299

#authors Jennings, C.G., Dyer, S.M., Burden, S.J., Proc. Natl. Acad. Sci. U.S.A. (1993) 90:2895-2899

#journal Muscle-specific ror-related receptor with a kringe domain defines a distinct class of receptor tyrosine kinases.

#title #defines a distinct class of receptor tyrosine kinases.

#cross-references MUID:93219391

Accession A47299

##status Preliminary

##molecule_type mRNA

##residues 1-946 #label JEN

##cross-references GB:M11311; NID:9290857; PIDN:AAA49285.1; PID:9290858

##experimental_source electric organ

##note sequence extracted from NCBI backbone (NCBIN:128724, NCBI:P128726)

CLASSIFICATION #superfamily Torpedo ror-related receptor; kringe homology; protein kinase homology

KEYWORDS ATP; glycoprotein; kringe; phosphotransferase; transmembrane protein; tyrosine-specific protein kinase

FEATURE

464-542 #domain kringe homology #label KR3\

572-588 #domain transmembrane #status predicted #label TMN\

650-940 #domain protein kinase homology #label KIN\

658-666 #region protein kinase ATP-binding motif\

225,340,477,544 #binding_site carbohydrate (asn) (covalent) #status predicted

SUMMARY #length 946 #molecular_weight 105892 #checksum 87117

Query Match 33.9%; Score 207; DB 1; Length 946;

Best Local Similarity 39.4%; Pred. No. 3,026-23;

Matches 26; Conservative 13; Mismatches 23; Indels 4; Gaps 4;

DB 476 VNTASISICORSEQAPHFHRRLPELPPEL-ANSDNFCRNPGESEPPWC-TYMDRP-I 532

QY 47 VSVTDFCAPCLRMALVEPFLERSPSPASWQDRGQRHNCRSPDAG-RPWCFTGDARGKV 105

333 RMEFCN 538

106 DWGYCD 111

ENTRY 3

TITLE KRFH12 #type complete

ALTERNATE_NAMES coagulation factor XIIa (EC 3.4.21.38) precursor - human

ORGANISM #formal_name Homo sapiens #common_name man

DATE 27-Nov-1985 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

ACCESSIONS A29411; A26814; A00930; A25191; A22248; A21037

REFERENCE A29411

#authors Cool, D.E.; MacGillivray, R.T.A.

#journal J. Biol. Chem. (1987) 262:13662-13673

#title Characterization of the human blood coagulation factor XII gene. Intron/exon gene organization and analysis of the 5'-flanking region.

#cross-references MUID:88007593

Accession A29411

##molecule_type DNA

##residues 1-615 #label COO

##cross-references GB:M17466; GB:J02807; NID:9180355; PIDN:AA59490.1; PID:9180357

REFERENCE A26814

#authors Tripodi, M.; Citarella, F.; Guida, S.; Galeffi, P.; Fantoni, A.; Correse, R.

#journal Nucleic Acids Res. (1986) 14:3146

#title cDNA sequence coding for human coagulation factor XII (Hageman).

#cross-references MUID:86176794

Accession A26814

##molecule_type mRNA

##residues 4-615 #label TRI

##cross-references GB:M13135; NID:9182291; PIDN:AAA70225.1; PID:9182292

REFERENCE A00930

#authors Cool, D.E.; Edgell, C.J.S.; Louie, G.V.; Zoller, M.J.; Brayer, G.D.; MacGillivray, R.T.A.

#journal J. Biol. Chem. (1985) 260:13666-13676

#title Characterization of human blood coagulation factor XII cDNA. Prediction of the primary structure of factor XII and the tertiary structure of beta-factor XIIa.

#cross-references MUID:86033830

Accession A00930

##molecule_type mRNA

##residues 14-332, 'S', 334-615 #label CO2

##cross-references GB:M11723; NID:9180358; PIDN:AAA51986.1; PID:9180359

REFERENCE A25191

#authors Que, B.G.; Davie, E.W.

#journal Biochemistry (1986) 25:1525-1528

#title Characterization of a cDNA coding for human factor XII (Hageman factor).

#cross-references MUID:86216049

Accession A25191

##molecule_type mRNA

##residues 146-378, 'G', 380-615 #label OUE

##cross-references GB:M3147; NID:9180360; PIDN:AAA70224.1; PID:9180361

REFERENCE A22248

#authors McMullen, B.A.; Fujikawa, K.

#journal J. Biol. Chem. (1985) 260:5328-5341

#title Amino acid sequence of the heavy chain of human alpha-factor XIIa (activated Hageman factor).

#cross-references MUID:85182674

Accession A22248

##molecule_type protein

##residues 20-379 #label MCM

REFERENCE A21037

#authors Fujikawa, K.; McMullen, B.A.

#journal J. Biol. Chem. (1985) 258:10924-10933

#title Amino acid sequence of human beta-factor XIIa.

#cross-references MUID:83291041

Accession A21037

##molecule_type protein

##residues 354-362; 373-615 #label FUJ

REFERENCE A44606

#authors Harris, R.J.; Ling, V.T.; Spellman, M.W.

#journal J. Biol. Chem. (1992) 267:5102-5107

#title O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C.

#cross-references MUID:92184750

Accession A44606

##molecule_type protein

##residues 354-362; 373-615 #label FUJ

GENETICS

#gene GDB:FL2

#cross-references GDB:119892; OMIM:234000

map_position 5q34-5qter

#introns 19/3; 39/1; 72/2; 96/1; 133/1; 177/1; 212/1; 267/2; 340/1; 417/2; 463/1; 511/1; 560/3

COMPLEX

#description factor XIIa catalyzes the proteolytic activation of plasminogen, plasma prekallikrein, and coagulation factors VII and IX; factor XII can catalyze the proteolytic activation of prekallikrein

FUNCTION

#description factor XIIa catalyzes the proteolytic activation of plasminogen, plasma prekallikrein, and coagulation factors VII and IX; factor XII can catalyze the proteolytic activation of prekallikrein

CLASSIFICATION

#pathway blood coagulation; fibrinolysis

#superfamily coagulation factor XII; EGF homology; fibronectin type I repeat homology; fibronectin type II repeat homology; kringe homology; trypsin homology


```

##molecule-type DNA
##residues 1-562 ##label NYT
##cross-references GB:L00141
##note the codon given for residue 93 (ACC) is inconsistent
      with the authors' translation
REFERENCE
#A23529
#authors Frieze Degen, S.J.; Rajput, B.; Reich, E.
#journal J. Biol. Chem. (1986) 261:6972-6985
#title The human tissue plasminogen activator gene.
#cross-references MIM:86196143
#accession A23529
##molecule-type DNA
##residues 1-562 ##label DEG
##cross-references GB:R03021; NID:q339817; PIDN:AAA9809.1; PID:q339818
#J0562
#authors Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.
#journal Agric. Biol. Chem. (1991) 55:1225-1232
#title Purification and characterization of tissue plasminogen
      activator secreted by human embryonic lung diploid
      fibroblasts, IMR-90 cells.
#cross-references MIM:91291340
#accession J0562
##molecule-type mRNA
##residues 31-562 ##label ITA
##cross-references DDBJ:D01096; NID:q220128; PIDN:BA00881.1;
      PID:d1001345; PID:q441174
##experimental_source embryonic lung fibroblast IMR-90 cells
##note mature protein, was confirmed by protein sequencing
REFERENCE
#A93293
#authors Penlice, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar,
      G.A.; Ward, C.A.; Bennett, W.F.; Yelverton, E.; Seeburg,
      P.H.; Heyneker, H.L.; Goeddel, D.V.; Collen, D.
#journal Nature (1983) 301:214-221
#title Cloning and expression of human tissue-type plasminogen
      activator cDNA in Escherichia coli.
#cross-references MIM:83115262
#accession A93293
##molecule-type mRNA
##residues 1-562 ##label PEN
##cross-references GB:L00141
#S02125
##experimental_source melanoma cells
REFERENCE
#S02125
#authors Sasaki, H.; Saito, Y.; Hayashi, M.; Otsuka, K.; Niwa, M.
#journal Nucleic Acids Res. (1988) 16:5695
#title Nucleotide sequence of the tissue-type plasminogen activator
      cDNA from human fetal lung cells.
#cross-references MIM:88262579
#accession S02125
##status translation not shown
##molecule-type mRNA
##residues 1-562 ##label SAS
##cross-references EMBL:X07393; NID:q37243; PIDN:CA030302.1; PID:q37244
##experimental_source fetal lung cells
REFERENCE
#A91343
#authors Kagitani, H.; Tagawa, M.; Hatanaka, K.; Ikari, T.; Saito, A.;
      Bando, H.; Okada, K.; Matsuo, O.
#journal FEBS Lett. (1985) 189:145-149
#title Expression in Escherichia coli of finger-domain lacking
      tissue-type plasminogen activator with high fibrin
      affinity.
#cross-references MIM:85285620
#accession A91343
##molecule-type mRNA
##residues 1-38, 'G', 86-433, 'E', 435-562 ##label KAG
##experimental_source Detroit 562 cells; ATCC 138
REFERENCE
#A93951
#authors Edlund, T.; Ny, T.; Ranby, M.; Hedén, L.O.; Palm, G.;
      Holmgren, E.; Josephson, S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1983) 80:349-352
#title Isolation of cDNA sequences coding for a part of human tissue
      plasminogen activator.
#cross-references MIM:83169656
#accession A93951
##molecule-type mRNA
##residues 251-358 ##label EDL
##experimental_source melanoma cells
REFERENCE
#A90488
#authors Pohl, G.; Kallstrom, M.; Bergsdorf, N.; Wallen, P.; Jorvall, H.
#journal Biochemistry (1984) 23:3701-3707
#title Tissue plasminogen activator: peptide analyses confirm an
      indirectly derived amino acid sequence, identify the active
      site serine residue, establish glycosylation sites, and
      localize variant differences.
#cross-references MIM:85000468
#accession A91322
#authors Pohl, G.; Kaplan, L.; Einarsson, M.; Wallen, P.; Jorvall, H.
#journal FEBS Lett. (1984) 168:29-32
#title Differences between uterine and melanoma forms of tissue
      plasminogen activator.
#cross-references MIM:84158956
#accession A91322
##molecule-type protein
##residues 33-45;311-320 ##label POH
##experimental_source uterus
##note in the uterus, cleavage of the activation peptide may
      also occur after 38-Gln
REFERENCE
#A37567
#authors van Zonneveld, A.J.; Veerman, H.; Pannekoek, H.
#journal J. Biol. Chem. (1986) 261:14214-14218
#title annotation; fibrin binding site
#cross-references MIM:87033611
#accession A37568
#authors Verheijen, J.H.; Caspers, M.P.M.; Chang, G.T.G.; de Munck,
      G.A.W.; Pouwels, P.H.; Enger-Valk, B.E.
#journal EMBO J. (1986) 5:3525-3530
#title Involvement of finger domain and kringle 2 domain of
      tissue-type plasminogen activator in fibrin binding and
      stimulation of activity by fibrin.
#cross-references MIM:87161761
#accession A60902
#contents annotation; fibrin binding site
#cross-references MIM:89044681
#accession A54645
#contents annotation; novel forms of expressed recombinant t-Pa
REFERENCE
#A54645
#authors Harris, T.J.R.; Patel, T.; Marston, F.A.O.; Little, S.;
      Emtage, J.S.; Odenakker, G.; Volckert, G.; Romants, W.;
      Billiau, A.; De Somer, P.
#journal Mol. Biol. Med. (1986) 3:279-292
#title Cloning of cDNA coding for human tissue-type plasminogen
      activator and its expression in Escherichia coli.
#cross-references MIM:86284200
#accession A54645
##molecule-type mRNA
##residues 1-562 ##label HAR
##cross-references GB:M15518; NID:q190031; PIDN:AAA6011.1; PID:q190032
##note parts of this sequence were confirmed by peptide
      sequencing
REFERENCE
#I60110
#authors Reddy, V.B.; Garramone, A.J.; Sasak, H.; Wei, C.
#journal DNA (1987) 6:461-472
#title Expression of human uterine tissue-type plasminogen activator
      in mouse cells using BPV vectors.
#cross-references MIM:88054470
#accession I60110
#status translated from GB/EMBL/DBJ
##molecule-type mRNA

```


4 / VSVIDGAPCLRWAEVPPFLERSPPASVWQLRGQRHNFCRSPDGAGRPMCEYGDARGKYD 106

ORGANISM	#Journal_name	homo sapiens	#common_name	man
DATE	15-Oct-1995	#sequence_revision	16-Aug-1996	#text_change
	16-Jul-1999			

```
DATE      15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change
          16-Jul-1999
```



```

30-308      #product t-plasminogen activator chain A #status
38-75      #predicted #label ACH\
83-116      #domain fibronectin type I repeat homology #label 1F1\
124-205      #domain EGF homology #label EGF\
213-294      #domain kringle homology #label KR1\
309-553      #product t-plasminogen activator chain B #status
38-68,66-75,83-94, #predicted #label BCH\
88-105,107-116, #domain trypsin homology #label TRY\
124-205,145-187,
176-200,213-284,
234-276,265-289,
297-428,340-356,
348-417,442-516,
474-490,506-534,
149,481      #disulfide bonds #status predicted\
#binding site carbohydrate (asn) (covalent) #status
#predicted\
#cleavage_site Arg-Ile (plasmin, trypsin) #status
#predicted\
#active_site His, Asp, Ser #status predicted
SUMMARY      #length 559 #molecular-weight 62903 #checksum 8573
55,404,510
Query Match      26.8%: Score 164; DB 1; Length 559;
Best Local Similarity 41.0%: Pred. No. 8,98e-15;
Matches 25: Conservative 12; Mismatches 19; Indels 5; Gaps 4;
DB 232 ASCLPWNMT-LIGITYAWRANSQALGLGRNYCRNPDGDAKPCWCHMKDR-KLTWEYC 289
QY 54 APCLRMAEVPPELERSPPASWA--QLRGO-RHNFCRSPDGAGRPPCFYGDARGKVDWGXC 110
DB 290 D 290
QY 111 D 111

```

Search completed: Mon Mar 13 10:22:13 2000
 Job time : 10 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:21:05 2000; Maspar time 5.46 Seconds
 Tubular output not generated. 399,446 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (40-112) from US09147947A.pep (2 of 6)
 Perfect Score: 611
 Sequence: 1 CPAGEPWSVTDFGAPCLRW.....RPMCFYGDARGKYDMGYCDC 73

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: Swiss-Prot
 Triswissprot

Statistics: Mean 37.173; Variance 60.386; scale 0.616

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description	Pred. No.
1	611	100.0	875	1	NETR_HUMAN	1.14e-123
2	507	83.0	761	1	NETR_MOUSE	1.24e-97
3	185	30.3	615	1	FA12_HUMAN	4.54e-21
4	179	29.3	552	1	UROT_MOUSE	8.93e-20
5	179	29.3	562	1	UROT_HUMAN	8.93e-20
6	175	28.6	716	1	HGFL_MOUSE	6.42e-19
7	174	28.6	812	1	PLMN_BOVIN	1.05e-18
8	172	28.2	812	1	PLMN_MOUSE	2.80e-18
9	171	28.0	603	1	FA12_CAVPO	4.57e-18
10	171	28.0	790	1	PLMN_PTG	4.57e-18
11	169	27.7	593	1	FA12_BOVIN	1.21e-17
12	164	26.8	559	1	UROT_RAT	1.38e-16
13	163	26.7	728	1	HGF_HUMAN	2.23e-16
14	161	26.4	711	1	HGFL_HUMAN	5.86e-16
15	161	26.4	728	1	HGF_MOUSE	5.86e-16
16	161	26.4	728	1	HGF_RAT	5.86e-16
17	161	26.4	810	1	PLMN_ERIEU	5.86e-16
18	157	25.7	4548	1	APOL_HUMAN	3.99e-15
19	156	25.5	169	1	PLMN_RAT	6.43e-15
20	154	25.2	1420	1	APOL_MOUSE	1.67e-14
21	150	24.5	655	1	HGFL_HUMAN	1.11e-13
22	149	24.4	810	1	PLMN_MOUSE	1.78e-13
23	147	24.1	810	1	PLMN_HUMAN	4.54e-13

24 145 23.7 566 1 UROT_BOVIN TISSUE PLASMINOGEN ACT 1.16e-12
 25 142 23.2 432 1 UROK_RAT UROKINASE-TYPE PLASMIN 4.68e-12
 26 140 22.9 433 1 UROK_MOUSE UROKINASE-TYPE PLASMIN 1.18e-11
 27 139 22.7 394 1 UROK_DESRO SALIVARY PLASMINOGEN A 1.87e-11
 28 138 22.6 431 1 UROK_HUMAN UROKINASE-TYPE PLASMIN 2.96e-11
 29 138 22.6 442 1 UROK_PIG UROKINASE-TYPE PLASMIN 2.96e-11
 30 134 21.9 433 1 UROK_BOVIN UROKINASE-TYPE PLASMIN 1.84e-10
 31 132 21.6 622 1 THRB_HUMAN PROTHROMBIN PRECURSOR 4.57e-10
 32 131 21.4 618 1 THRB_MOUSE PROTHROMBIN PRECURSOR 7.18e-10
 33 129 21.1 433 1 UROK_PAPCY UROKINASE-TYPE PLASMIN 1.77e-09
 34 127 20.8 333 1 PLMN_CANFA PLASMINOGEN (EC 3.4.21 4.32e-09
 35 126 20.6 617 1 THRB_RAT PROTHROMBIN PRECURSOR 6.75e-09
 36 125 20.5 343 1 PLMN_SHEEP PLASMINOGEN (EC 3.4.21 1.05e-08
 37 125 20.5 625 1 PLMN_BOVIN PROTHROMBIN PRECURSOR 1.05e-08
 38 121 19.8 431 1 URT2_DESRO SALIVARY PLASMINOGEN A 6.14e-08
 39 121 19.8 477 1 URT2_DESRO SALIVARY PLASMINOGEN A 6.14e-08
 40 120 19.6 434 1 UROK_CHICK UROKINASE-TYPE PLASMIN 9.50e-08
 41 114 18.7 338 1 PLMN_HORSE PLASMINOGEN (EC 3.4.21 1.27e-06
 42 114 18.7 477 1 URT1_DESRO SALIVARY PLASMINOGEN A 1.27e-06
 43 93 15.2 325 1 PLMN_PETMA PLASMINOGEN (EC 3.4.21 6.71e-03
 44 92 15.1 284 1 YPV2_METTE HYPOTHETICAL 33.2 KD P 9.86e-03
 45 90 14.7 501 1 PWP1_HUMAN PERIODIC TRYPHOPHAN PR 2.12e-02

ALIGNMENTS

RESULT 1
 ID NETR_HUMAN STANDARD: PRT: 875 AA.
 AC P56730.15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).
 GN PRS12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 NC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RC [1]
 RP SEQUENCE FROM N.A.
 RN TISSUE-BRAIN:
 RX MEDLINE: 98201705.
 RA PROBA K., GSCHEWEND T.P., SONDEREGGER P.;
 RT "Cloning and sequencing of the cDNA encoding human neutrotypin";
 RL Biochem. Biophys. Acta 1396:143-147(1998).
 CC -!- FUNCTION: PLAYS A ROLE IN NEUROBLASTICITY AND THE PROTEOLYTIC
 ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
 LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 SRCR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb-sdb.ch).
 CC EMBL: AJ001531; CAA04816.1; -
 DR PROSITE: PS00134; TRYPSIN_HIS. 1.
 DR PROSITE: PS00135; TRYPSIN_SER. 1.
 DR PROSITE: PS00420; SERPACT_RECEPTOR. 3.
 KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 875 NEUTROTYPIN.
 FT DOMAIN 23 92 PROLINE-RICH.
 FT DOMAIN 93 165 KRINGLE.
 FT DOMAIN 170 271 SRCR 1.
 FT DOMAIN 280 381 SRCR 2.
 FT DOMAIN 387 487 SRCR 3.

```

FT DOMAIN 500 601 SRCR 4. SERINE. PROTEASE.
FT DOMAIN 619 875
FT DOMAIN 619 630 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 67D5272B CRC32:

Query Match 100.0%; Score 611; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 1,14e-123;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 93 CPAGEPWSVTDFGAPCLRWAEVPPFLERSPPASMAQLRGQRHNCRSPDGRPMCFY 152
40 CPAGEPWSVTDFGAPCLRWAEVPPFLERSPPASMAQLRGQRHNCRSPDGRPMCFY 99
153 DARGKVDWGYCDC 165
100 DARGKVDWGYCDC 112

RESULT 2
ID NEUR_MOUSE STANDARD: PRT; 761 AA.
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEUROTRYPsin, a novel multidomain serine protease expressed in the
DE PROTEASE 3 (BSSP-3).
DE PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GOSCHWEND T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neurotrypsin, a novel multidomain serine protease expressed in the
RT nervous system."
RT Mol. Cell. Neurosci. 9:207-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSUBOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
RT single-like structure and three scavenger receptor cysteine-rich
RT motifs."
RN [3]
RP Biochem. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
CC LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
CC AND AMYGALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: Y13192; CAA73646.1;
DR EMBL: D89871; BAA23986.1;
DR MGd; MG1:1100881; PRSS12.
DR PFAM: PF00530; SRCR; 3.
DR PFAM: PF00089; trypsin; 1.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.
KW Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 761 NEUROTRYPsin.
FT DOMAIN 85 157 POTENTIAL.
FT DOMAIN 156 267 KRINGLE.
FT DOMAIN 273 373 SRCR 1.
FT DOMAIN 386 487 SRCR 2.
FT DOMAIN 505 761 SRCR 3.
FT DOMAIN 505 761 SERINE. PROTEASE.
FT ACT_SITE 516 516 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 516 517 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 562 562 CHARGE RELAY SYSTEM.
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 POTENTIAL.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA: 84118 MW: 3F3CAF35 CRC32:

Query Match 83.0%; Score 507; DB 1; Length 761;
Best Local Similarity 78.1%; Pred. No. 1,24e-97;
Matches 57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

Db 85 CGAGESMGNAINLGVPCILHWDEVPPFLERSPPASMAELRGQRHNCRSPDGRPMCFY 144
40 CPAGEPWSVTDFGAPCLRWAEVPPFLERSPPASMAQLRGQRHNCRSPDGRPMCFY 99
145 NAQKVDWGYCDC 157
100 DARGKVDWGYCDC 112

RESULT 3
ID FA12_HUMAN STANDARD: PRT; 615 AA.
AC P00748;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COAGULATION FACTOR XII PRECURSOR (EC 3.4.21.38) (HAGEMAN FACTOR)
DE (HAF).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88007593.
RA COOL D.E., MCGILLIVRAY R.T.A.;
RT "Characterization of the human blood coagulation factor XII gene.
RT Intron/exon gene organization and analysis of the 5'-flanking
RT region."
RN [2]
RP J. Biol. Chem. 262:13662-13673(1987).
RN [3]
RP SEQUENCE OF 4-615 FROM N.A.
RX MEDLINE: 86176794.
RA TRIPODI M., CITARELLA F., GUIDA S., GALEFFI P., FANTONI A.,
RA CORPSE R.;
RT "cDNA sequence coding for human coagulation factor XII (Hageman).";
RT Nucleic Acids Res. 14:3146-3146(1986).
RN [3]
RP SEQUENCE OF 14-615 FROM N.A.
RX MEDLINE: 86033830.
RA COOL D.E., ENGELL C.J.S., LOUIE G.V., ZOLLER M.J., BAYER G.D.,
RA MCGILLIVRAY R.T.A.;
RT "Characterization of human blood coagulation factor XII cDNA.

```


RX MEDLINE: 88262579.
 RA SASAKI H., SAITO Y., HAYASHI M., OTSUKA K., NIMA M.;
 RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA
 from human fetal lung cells.";
 RL Nucleic Acids Res. 16:5695-5695(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 88054470.
 RA REDDY V.B., GARRAMONE A.J., SASAK H., WEI C.M., WATKINS P., GALLI J.,
 RA HSUNG N.;
 RT "Expression of human uterine tissue-type plasminogen activator in
 mouse cells using BPV vectors.";
 RL DNA 6:461-472(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 86196143.
 RA DEGEN S.J.F., RAJPUT B., REICH E.;
 RT "The human tissue plasminogen activator gene.";
 RL J. Biol. Chem. 261:6972-6985(1986).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 84298137.
 RA NY T., ELGH F., LUND B.;
 RT "The structure of the human tissue-type plasminogen activator gene:
 correlation of intron and exon structures to functional and
 structural domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).
 RN [6]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 86284200.
 RA HARRIS T.J., PATEL T., MARSTON F.A., LITTLE S., EMRAGE J.S.,
 RA OPDENAKKER G., VOLCKERT G., ROMBAUTS W., BILLIAU A., SOMER P.;
 RT "Cloning of cDNA coding for human tissue-type plasminogen activator
 and its expression in Escherichia coli.";
 RL Mol. Biol. Med. 3:279-292(1986).
 RN [7]
 RP SEQUENCE OF 251-358 FROM N.A.
 RA MEDLINE: 83169656.
 RA EDLUND T., NY T., RANBY M., HEDEN L.-O., PALM G., HOLMGREN E.,
 RA JOSEPHSON S.;
 RT "Isolation of cDNA sequences coding for a part of human tissue
 plasminogen activator.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352(1983).
 RN [8]
 RP SEQUENCE OF 31-562 FROM N.A.
 RA MEDLINE: 91291340.
 RA ITAKAKI Y., YASUDA H., MORINAGA T., MITSUDA S., HIGASHIO K.;
 RT "Purification and characterization of tissue plasminogen activator
 secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";
 RL Agric. Biol. Chem. 55:1225-1232(1991).
 RN [9]
 RP SEQUENCE OF 36-562.
 RA MEDLINE: 85000468.
 RA POHL G., KALLISTROM M., BERGSDORF N., WALLEN P., JOERNVALL H.;
 RT "Tissue plasminogen activator: peptide analyses confirm an indirectly
 derived amino acid sequence, identify the active site serine residue,
 establish glycosylation sites, and localize variant differences.";
 RL Biochemistry 23:3701-3707(1984).
 RN [10]
 RP SEQUENCE FROM N.A. (SMALL ISOFORM).
 RA TISSUE-OMBITICAL VEIN.
 RA MEDLINE: 90192129.
 RA SIEBERT P.D., FONG K.;
 RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from
 human endothelial cells.";
 RL Nucleic Acids Res. 18:1086-1086(1990).
 RN [11]
 RP STRUCTURE OF CARBOHYDRATES.
 RA MEDLINE: 90092112.
 RA PEEFERER G., SCHMIDT M., STRUBE K.-H., GEYER R.;
 RT "Carbohydrate structure of recombinant human uterine tissue
 plasminogen activator expressed in mouse epithelial cells.";
 RL Eur. J. Biochem. 186:273-286(1989).

RN [12]
 RP DISULFIDE BONDS IN KRINGLE 2.
 RA MEDLINE: 91244765.
 RA VLACHOS C.J., WILHELM O.G., HASSELL T., JASKUNAS S.R., BANG N.U.;
 RT "Disulfide pairing of the recombinant kringle-2 domain of tissue
 plasminogen activator produced in Escherichia coli.";
 RL J. Biol. Chem. 266:10070-10072(1991).
 RN [13]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.
 RA MEDLINE: 96200985.
 RA LABA D., BAUER M., HUBER R., FISCHER S., RUDDOLPH R., KOHNERT U.,
 RA BODE M.;
 RT "The 2.3 A crystal structure of the catalytic domain of recombinant
 two-chain human tissue-type plasminogen activator.";
 RL J. Mol. Biol. 258:117-135(1996).
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.
 RA MEDLINE: 92118803.
 RA DE VOS A., ULTSCH M.H., KELLEY R.F., PADMANABHAN K., TULINSKI A.,
 RA WESTBROOK M.L., KOSSIAKOF A.A.;
 RT "Crystal structure of the kringle 2 domain of tissue plasminogen
 activator at 2.4-A resolution.";
 RL Biochemistry 31:270-279(1992).
 RN [15]
 RP STRUCTURE BY NMR OF KRINGLE 2.
 RA MEDLINE: 90122799.
 RA BYEON I.-J.L., KELLEY R.F., LILINAS M.;
 RT "1H NMR structural characterization of a recombinant kringle 2 domain
 from human tissue-type plasminogen activator.";
 RL Biochemistry 28:9350-9360(1989).
 RN [16]
 RP STRUCTURE BY NMR OF KRINGLE 2.
 RA MEDLINE: 91200042.
 RA BYEON I.-J.L., KELLEY R.F., LILINAS M.;
 RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR
 assignments and secondary structure.";
 RL Eur. J. Biochem. 197:155-165(1991).
 RN [17]
 RP STRUCTURE BY NMR OF KRINGLE 2.
 RA MEDLINE: 92106329.
 RA BYEON I.-J.L., LILINAS M.;
 RT "Solution structure of the tissue-type plasminogen activator kringle
 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic
 drug.";
 RL J. Mol. Biol. 222:1035-1051(1991).
 RN [18]
 RP STRUCTURE BY NMR OF 38-85.
 RA MEDLINE: 92292163.
 RA DOWNING A.K., DRISCOLL P.C., HARVEY T.S., DUDGEON T.J., SMITH B.O.,
 RA BARON M., CAMPBELL I.D.;
 RT "Solution structure of the fibrin binding finger domain of
 tissue-type plasminogen activator determined by 1H nuclear magnetic
 resonance.";
 RL J. Mol. Biol. 225:821-833(1992).
 RN [19]
 RP STRUCTURE BY NMR OF 36-126.
 RA MEDLINE: 96027104.
 RA SMITH B.O., DOWNING A.K., DRISCOLL P.C., DUDGEON T.J., CAMPBELL I.D.;
 RT "The solution structure and backbone dynamics of the fibronectin type
 I and epidermal growth factor-like pair of modules of tissue-type
 plasminogen activator.";
 RL Structure 3:823-833(1995).
 CC -1- FUNCTION: TPA CONVERTS PLASMINOGEN TO PLASMIN BY HYDROLYZING A
 SINGLE R-Y BOND IN PLASMINOGEN. ACTIVE IN TISSUE REMODELING AND
 DESTRUCTION, PARTICULARLY IN FIBRINOLYSIS, AND IN CELL MIGRATION.
 CC -1- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-1-VAL BOND IN
 PLASMINOGEN TO FORM PLASMIN.
 CC -1- SUBUNIT: CLEAVAGE AFTER ARG-310 BY PLASMIN OR TRYPSIN RESULTS IN
 A TWO-CHAIN FORM OF THE MOLECULE. THESE TWO HALVES ARE HELD
 TOGETHER BY ONE OR MORE DISULFIDE BONDS.
 CC -1- ALTERNATIVE PRODUCTS: THERE EXISTS A SHORT VARIANT PROBABLY DUE TO
 ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- MISCELLANEOUS: TPA ATTACHES TO THE KRINGLE STRUCTURE OF THE A

Note: remainder of annotations omitted.

Matches	25;	Conservative	15;	Mismatches	23;	Indels	3;	Gaps	3;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

48 SVTDFGAPCLRWAEPFLEL-SPASWAQLRGQ-RHNFCSRSPDGAGRPWCIFYGDARGKV 105

106 DWGYCD 111

HGFL_MOUSE STANDARD; PRT; 716 AA

STIMULATORY PROTEIN) (MSP).

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

STRAIN-BALB/C; TISSUE-LIVER;

growth factor-like protein: expression during development.";

CONSERVED.

DEVELOPMENTAL STAGE: IS EXPRESSED AT LOW LEVELS DURING GESTATION JUST BEFORE BIRTH THE LEVEL INCREASES DRAMATICALLY AND REMAINS

or send an email to license@isb-sib.ch)

HSSP; P00747; 1PMK.

PFAM; PF00051; kringler; 4.

CHAIN	32	716	H
-------	----	-----	---

DOMAIN	292	370	KIN
--------	-----	-----	-----

DISULFID	60	66	BY
DISULFID	50	70	BY

DISULFID	191	268	B
DISULFID	157	181	B

DISULFID	292	370	B
DISULFID	240	263	B

DISULFID	3 / 9	45 /	B
DISULFID	400	440	B

DISULEFD	512	528	B
DISULFID	607	672	B

CARBOHYD	72	.72	P6
CARBOHYD	173	1.73	P6

CONFLICT	19	19	P
STORY	21	20	P

Local Similarity 35.4%; Pre

—

RESULT 7
ID PLMN_BOVIN STANDARD: PRT: 812 AA.
AC P06868; Q28162;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7).
GN PLG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN (1)
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC BERGLUND L., ANDERSEN M.D., PETERSEN T.E.;
RC Cloning and characterization of the bovine plasminogen cDNA.";
RC Int. Dairy J. 5:593-603(1995).
[2]
[2] SEQUENCE OF 27-812, AND CARBOHYDRATE-BINDING SITES.
RX MEDLINE: 85203906.
RA SCHALLER J., MOSER P.W., DANNEGER-MULLER G.A.K., ROSSELET S.J.,
RA KAMFER U., RICKLI E.E.;
RA "Complete amino acid sequence of bovine plasminogen. Comparison with
RA human plasminogen.";
RA Eur. J. Biochem. 149:267-278(1985).
RN [3]
RN SEQUENCE OF 706-812 FROM N.A.
RX MEDLINE: 85023311.
RA MALINOWSKI D.P., SADLER J.E., DAVIE E.W.;
RA "Characterization of a complementary deoxyribonucleic acid coding for
RA human and bovine plasminogen.";
RA Biochemistry 23:4243-4250(1984).
CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION. IT WEAKENS THE WALLS OF THE
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS. BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79402; CAA55939.1; -
DR EMBL: K02935; AAA30714.1; -
DR PIR: A25835; PLBO.
DR HSSP: P00747; 2PK4.
DR PROSITE: PS00021; KRINGLE_1; 5.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PROSITE: PS00700; KRINGLE_2; 5.
DR PFAM: PF00051; kringle; 5.
DR PFAM: PF00089; trypsin; 1.
KW Hydrolase; Serine protease; Plasma; Glycoprotein; Fibrinolysis;
KW Tissue remodeling; Blood coagulation; Kringle; Zymogen; Signal.
FT SIGNAL 1 26
FT CHAIN 27 812 PLASMINOGEN.
FT CHAIN 27 583 HEAVY CHAIN A.

FT CHAIN 584 812 LIGHT CHAIN B.
FT DOMAIN 584 812 CATALYTIC.
FT DOMAIN 110 188 KRINGLE 1.
FT DOMAIN 192 269 KRINGLE 2.
FT DOMAIN 282 359 KRINGLE 3.
FT DOMAIN 384 461 KRINGLE 4.
FT DOMAIN 485 564 KRINGLE 5.
FT CARBOHYD 315 315
FT CARBOHYD 365 365
FT ACT_SITE 624 624 CHARGE RELAY SYSTEM.
FT ACT_SITE 667 667 CHARGE RELAY SYSTEM.
FT ACT_SITE 762 762 CHARGE RELAY SYSTEM.
FT CONFLICT 335 335 N -> D (IN REF. 2).
FT CONFLICT 516 516 Q -> H (IN REF. 2).
FT CONFLICT 555 555 P -> L (IN REF. 2).
FT CONFLICT 744 744 T -> R (IN REF. 3).
SQ SEQUENCE 812 AA; 91216 MW; 3E3C0328 CRC32;
Query Match 28.5%; Score 174; DB 1; Length 812;
Best Local Similarity 35.9%; Pred. No. 1,05e-18;
Matches 23; Conservative 11; Mismatches 27; Indels 3; Gaps 3;
DB 294 VAYTESGTCQMRSEOTPHKHNRTPEFPC-KNLENYGRNPGKAPKC-YT-TNSEVR 350
QY 47 VSVTDGAPCLMAEVPPELERSPPASWALRGQRHNFCSBDGAGRPWCFTGDARKVD 106
DB 351 WEYC 354
QY 107 WGYC 110
RESULT 8
ID PLMN_MOUSE STANDARD: PRT: 812 AA.
AC P20918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7) [CONTAINS: ANGIOSTATIN].
GN PLG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 91184812.
RA DEGEN S.J., BELL S.M., SCHAEFER L.A., ELLIOTT R.W.;
RA "Characterization of the cDNA coding for mouse plasminogen and
RA localization of the gene to mouse chromosome 17.";
RA Genomics 8:49-61(1990).
RN [2]
RN CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE: 95042728.
RA O'REILLY M.S., HOLMGREN L., SHING Y., CHEN C., ROSENTHAL R.A.,
RA MOSS M., LANE W.S., CAO Y., SAGE E.H., FOLKMAN J.;
RA "Angiostatin: a novel angiogenesis inhibitor that mediates the
RA suppression of metastases by a Lewis lung carcinoma.";
RA Cell 79:315-328(1994).
CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION. IT WEAKENS THE WALLS OF THE
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- FUNCTION: ANGIOSTATIN IS AN ANGIOGENESIS INHIBITOR THAT BLOCKS
CC NEOVASCULARIZATION AND GROWTH OF EXPERIMENTAL PRIMARY AND
CC METASTATIC TUMORS IN VIVO.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS. BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.


```
FT NON_TER 1 18
FT SIGNAL 19 358
FT CHAIN 19 358
FT CHAIN 399 603
FT DOMAIN 46 87
FT DOMAIN 93 130
FT DOMAIN 132 172
FT DOMAIN 173 209
FT DOMAIN 216 294
FT DOMAIN 312 342
FT DOMAIN 359 603
FT ACT_SITE 398 398
FT ACT_SITE 447 447
FT ACT_SITE 551 551
FT ACT_SITE 551 551
FT DISULFID 97 109
FT DISULFID 103 118
FT DISULFID 120 129
FT DISULFID 134 162
FT DISULFID 160 169
FT DISULFID 177 188
FT DISULFID 182 197
FT DISULFID 199 208
FT DISULFID 216 294
FT DISULFID 237 276
FT DISULFID 265 289
FT DISULFID 345 472
FT DISULFID 383 399
FT DISULFID 391 461
FT DISULFID 422 425
FT DISULFID 488 557
FT DISULFID 520 536
FT DISULFID 547 578
FT CARBOHYD 248 248
FT CARBOHYD 270 270
FT CARBOHYD 419 419
SQ SEQUENCE 603 AA: 66795 MW: 35C16971 CRC32:
Query Match 28.0%; Score 171; DB 1; Length 603;
Best Local Similarity 37.3%; Pred. No. 4.57e-18;
Matches 22; Conservative 12; Mismatches 23; Indels 2; Gaps 2;
```

```
Db 234 GAKCORMASEATYRNMTAQLRGLGH-HTCRNPDNTPRCWCFVWNG-NRLSMETCD 230
| | | | | : : : : : | : | | | | | | | | | : : | | | |
| | | | | : : : : : | : | | | | | | | | | : : | | | |
Qy 53 GACPLMAEVPFLERSPPASNAQLRGQRHNFCRSPDAGRPWCFYGDARGVDMGYCD 111
```

```
PLMN_PIG STANDARD: PRT: 790 AA.
P06867:
01-JAN-1988 (Rel. 06, Created)
```

```
01-FEB-1991 (Rel. 17, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
PLASMINOGEN (EC 3.4.21.7).
```

```
Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1]
```

```
SEQUENCE OF 1-560.
"Amnio acid sequence of the heavy chain of porcine plasmin. Comparison
of the carboxydrate attachment sites with the human and bovine
species.";
```

```
Fibrinolysis 1:91-102(1987).
```

```
[2]
```

```
SEQUENCE OF 450-790.
MEDLINE: 85203907.
```

```
MARTI T., SCHALLER J., RICKLI E.E.;
"Determination of the complete amino-acid sequence of porcine
miniplasminogen.";
```

```
Eur. J. Biochem. 149:279-285(1985).
-1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
EMERYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION.
```

```
CC AND INFLAMMATION. IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOMES, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- SIMILARITY: CONTRAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC PIR: A25834; A25834.
CC PIR: S03733; S03733.
CC HSP: P00747; SHPG.
CC PROSITE: PS00134; TRYPSIN_HIS: FALSE_NEG.
CC PROSITE: PS00021; KRINGLE_1; 5.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS50070; KRINGLE_2; 5.
CC PFAM: PF00051; kringle; 5.
CC PFAM: PF00089; trypsin; 1.
CC Hydrolase. Serine protease; plasma; glycoprotein; fibrinolysis;
CC Tissue remodeling; blood coagulation; kringle; zymogen.
CC CHAIN 1 560
CC CHAIN 561 790
CC DOMAIN 561 790
CC DOMAIN 84 162
CC DOMAIN 166 243
CC DOMAIN 256 333
CC DOMAIN 358 435
CC DOMAIN 461 540
CC ACT_SITE 602 602
CC ACT_SITE 645 645
CC ACT_SITE 740 740
CC CARBOHYD 249 249
SQ SEQUENCE 790 AA: EE597814 CRC32:
Query Match 28.0%; Score 171; DB 1; Length 790;
Best Local Similarity 37.5%; Pred. No. 4.57e-18;
Matches 24; Conservative 8; Mismatches 29; Indels 3; Gaps 3;
```

```
Db 268 VSVTASGHTQCRMSQSPFKHNRTPEFPC-KNLENYCRNPDGETAPWC-YTTD-SEVR 324
| | | | | : : : : : | : | | | | | | | | | : : | | | |
| | | | | : : : : : | : | | | | | | | | | : : | | | |
Qy 47 VSVTDEGAPCLMAEVPFLERSPPASNAQLRGQRHNFCRSPDAGRPWCFYGDARGKVD 106
```

```
Db 325 WDYC 328
| | |
Qy 107 WGYC 110
```

```
RESULT 11
ID FA12_BOVIN STANDARD: PRT: 593 AA.
```

```
AC P98140:
01-FEB-1996 (Rel. 33, Created)
```

```
01-FEB-1996 (Rel. 33, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
```

```
COAGULATION FACTOR XII PRECURSOR (EC 3.4.21.38) (HAGEMAN FACTOR)
DE (HAF) (FRAGMENT).
GN F12.
```

```
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae;
OC Bovinae; Bos.
```

```
[1]
```

```
SEQUENCE FROM N.A.
TISSUE-LIVER;
MEDLINE: 94242782.
```

```
SHIBUYA Y., SEMBA U., OKABE H., KAMBARA T., YAMAMOTO T.;
"Primary structure of bovine Hageman factor (blood coagulation factor
XII): comparison with human and guinea pig molecules.";
```

```
RT Biochim. Biophys. Acta 1206:63-70(1994).
```

```
[2]
```


RT identification of a gene coding for four kringle domains with
 CC homology to hepatocyte growth factor."
 RL Biochemistry 30:9768-9780(1991).
 CC -1- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
 CC CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
 CC CONSERVED.
 CC -1- PTM: MAY BE CLEAVED AFTER AA 484, TO YIELD A TWO-CHAIN MOLECULE
 CC HELD TOGETHER BY DISULFIDE BONDS, OR TWO SEPARATE POLYPEPTIDES.
 CC -1- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M74178; AAA50165.1; -
 CC EMBL: U37055; AAC50471.1; -
 CC PIR: A40331; A40331.
 CC HSSP: P00763; ISLM.
 CC MIM: 142408; -
 CC DR PROSITE: PS00021; KRINGLE_1; 4.
 CC DR PROSITE: PS50070; KRINGLE_2; 4.
 CC DR PFAM: PF00051; kringle; 4.
 CC DR PFAM: PF00089; trypsin; 1.
 CC KRingle: Glycoprotein; Serine protease homolog; Signal;
 CC Polymorphism.
 CC KW
 CC FT SIGNAL 1 31 POTENTIAL.
 CC FT CHAIN 32 711 HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
 CC FT DOMAIN 32 109 PAP.
 CC FT DOMAIN 110 186 KRINGLE 1.
 CC FT DOMAIN 191 268 KRINGLE 2.
 CC FT DOMAIN 283 361 KRINGLE 3.
 CC FT DOMAIN 370 448 KRINGLE 4.
 CC FT DOMAIN 484 711 SERINE PROTEASE-LIKE.
 CC FT DOMAIN 56 78
 CC FT DISULFID 60 66 BY SIMILARITY.
 CC FT DISULFID 110 186 BY SIMILARITY.
 CC FT DISULFID 131 169 BY SIMILARITY.
 CC FT DISULFID 157 181 BY SIMILARITY.
 CC FT DISULFID 191 268 BY SIMILARITY.
 CC FT DISULFID 194 324 INTERCHAIN (BY SIMILARITY).
 CC FT DISULFID 212 251 BY SIMILARITY.
 CC FT DISULFID 240 263 BY SIMILARITY.
 CC FT DISULFID 283 361 BY SIMILARITY.
 CC FT DISULFID 304 343 BY SIMILARITY.
 CC FT DISULFID 332 355 BY SIMILARITY.
 CC FT DISULFID 370 448 BY SIMILARITY.
 CC FT DISULFID 391 431 BY SIMILARITY.
 CC FT DISULFID 419 443 BY SIMILARITY.
 CC FT DISULFID 468 588 INTERCHAIN (BY SIMILARITY).
 CC FT DISULFID 507 523 BY SIMILARITY.
 CC FT DISULFID 602 667 BY SIMILARITY.
 CC FT DISULFID 632 646 BY SIMILARITY.
 CC FT DISULFID 657 685 BY SIMILARITY.
 CC FT CARBOHYD 72 72 POTENTIAL.
 CC FT CARBOHYD 296 296 POTENTIAL.
 CC FT CARBOHYD 615 615 POTENTIAL.
 CC FT VARIANT 13 13 Y->C.
 CC FT VARIANT 13 13 /FTID-VAR_006631.
 CC FT VARIANT 212 212 C->F.
 CC FT SEQUENCE 711 AA; 80379 MW; C9978F05 CRC32;
 CC SQ
 CC Query Match 26.4%; Score 161; DB 1; Length 711;
 CC Best Local Similarity 46.7%; Pred. No. 5.86e-16;
 CC Matches 21; Conservative 3; Mismatches 20; Indels 1; Gaps 1;

Qy 53 GAPLRWAEVPPFLERSPPASMAOLRGQRHNFCSRSPDGAGRPWCF 97
 ID HGF_MOUSE STANDARD; PRT; 728 AA.
 AC Q08048;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-JUN-1998 (Rel. 36, last annotation update)
 DE HEPATOCYTE GROWTH FACTOR PRECURSOR (SCATTER FACTOR) (SF)
 DE (HEPATOPOIETIN A).
 GN HGF.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 496-504.
 RC TISSUE=MAMMARY FIBROBLAST.
 RX MEDLINE: 94183257.
 RA SASAKI M., NISHIO M., SASAKI T., ENAMI J.;
 RT "Identification of mouse mammary fibroblast-derived mammary growth
 RT factor as hepatocyte growth factor."
 RL Biochem. Biophys. Res. Commun. 199;772-779(1994).
 CC -1- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
 CC HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPHIC FACTOR, AND ACTS
 CC AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.
 CC IT HAS NO DETECTABLE PROTEASE ACTIVITY.
 CC -1- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A
 CC DISULFIDE BOND.
 CC -1- ALTERNATIVE PRODUCTS: A SHORT FORM OF HGF IS PRODUCED BY
 CC ALTERNATIVE RNA SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF THE
 CC LONG FORM.
 CC -1- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10212; BAA01064.1; -
 CC EMBL: D10213; BAA01065.1; -
 CC DR HSSP: P14210; 2HGF.
 CC DR MGD; MGI:96079; HGF.
 CC DR PROSITE: PS00021; KRINGLE_1; 4.
 CC DR PROSITE: PS50070; KRINGLE_2; 4.
 CC DR PFAM: PF00051; kringle; 4.
 CC DR PFAM: PF00089; trypsin; 1.
 CC KW Growth factor; kringle; Glycoprotein; Serine protease homolog;
 CC Signal; Alternative splicing.
 CC FT SIGNAL 1 32
 CC FT CHAIN 496 728
 CC FT MOD_RES 33 33
 CC FT DOMAIN 33 128
 CC FT DOMAIN 129 207 KRINGLE 1.
 CC FT DOMAIN 212 289 KRINGLE 2.
 CC FT DOMAIN 306 384 KRINGLE 3.
 CC FT DOMAIN 392 470 KRINGLE 4.
 CC FT DOMAIN 496 728 SERINE PROTEASE-LIKE.
 CC FT DISULFID 71 97 BY SIMILARITY.
 CC FT DISULFID 75 85 BY SIMILARITY.
 CC FT DISULFID 488 607 INTERCHAIN (BY SIMILARITY).
 CC FT CARBOHYD 295 295 POTENTIAL.
 CC FT CARBOHYD 403 403 POTENTIAL.
 CC FT CARBOHYD 569 569 POTENTIAL.
 CC FT CARBOHYD 656 656 POTENTIAL.

FT	VARSPLIC	163	167	MISSING (IN SHORT ISOFORM)
SQ	SEQUENCE	728 AA;	82944 MN;	967CDFLE CRC32;

```

SQ      SEQUENCE      728 AA;  82944 MW;  967CDF1E CRC32;

```

Query Match	26.4%;	Score 161;	DB 1;	Length 728;
Best Local Similarity	36.1%;	Prod No	5	86e-16.

Query Match	26.48;	Score 161;	DB 1;	Length 728;
Best Local Similarity	36.18;	Pred. No. 5,866-16;		
Matches	22;	Conservative	10;	Mismatches 26;
			Indels 3;	Gaps 3;

Matches	22;	Conservative	10;	Mismatches	26;	Indels	3;	Gaps	3;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

Db 227 TEGSGTCQRMDDQTPHRAKFLPERYPD-KGFDDNYCRNPDGKPRPWCYTLDP-DT-PWEY 283

QY 50 TDFGAPCLRWAEVPPFLERSPPASMAQLRGQRHNFCSRSPDAGRPWCFTYGDARGKVDWGY 109

Db 284 C 284

QY 110 C 110

Search completed: Mon Mar 13 10:21:12 2000
Job time : 7 secs.

Search completed.
Job time : 7 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

App protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 13 10:21:30 2000; Maspar time 14.19 Seconds

Tabular output not generated. 356,626 Million cell updates/sec

Title: >US-09-147-947-6

Description: (40-112) from US09147947A.ppt (2 of 6)

Perfect Score: 611

Sequence: 1 CPAGEPWSVTDFGAPCLRM.....RPMCFYGDARGKVDWGYCDC 73

Scoring table: PAM 150

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database:

sptrembl12
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
 9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 37.029; Variance 62.983; scale 0.568

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
207	33.9	946	13	007153	RECEPTOR TYROSINE KINA	3.09e-24	
188	30.8	597	11	035727	FACTOR XII	3.03e-20	
173	28.3	716	11	P70521	HEPATOCTE GROWTH FACT	3.72e-17	
167	27.3	560	4	014520	HGF ACTIVATOR LIKE PRO	6.13e-16	
165	27.0	616	6	097507	FXII	1.55e-15	
163	26.7	290	4	002935	HEPATOCTE GROWTH FACT	3.92e-15	
163	26.7	296	4	014519	COMPETITIVE HGF ANTAGO	3.92e-15	
162	26.5	2869	6	028398	APOLIPROTEIN (FRAGME	6.22e-15	
161	26.4	567	4	013208	HEPATOCTE GROWTH FACT	9.86e-15	
161	26.4	711	4	014870	MACROPHAGE-STIMULATING	9.86e-15	
161	26.4	728	11	064007	HEPATOCTE GROWTH FACT	9.86e-15	
161	26.4	748	11	061652	HEPATOCTE GROWTH FACT	9.86e-15	
160	26.2	806	6	018783	PLASMINOGEN	1.56e-14	
159	26.0	710	13	091402	HEPATOCTE GROWTH FACT	2.47e-14	
152	24.9	454	6	046506	APOLIPROTEIN A (FRAG	6.02e-13	
150	24.5	717	13	P70006	HEPATOCTE GROWTH FACT	1.49e-12	
149	24.4	725	13	042341	HGF ALPHA-CHAIN (FRAGM	2.34e-12	
149	24.4	726	13	090978	HEPATOCTE GROWTH FACT	2.34e-12	
147	24.1	704	13	090865	HEPATOCTE GROWTH FACT	5.74e-12	
147	24.1	810	4	015146	PLASMINOGEN PRECURSOR	5.74e-12	

ALIGNMENTS

21	141	23.1	716	13	091691	GROWTH FACTOR LIVERIN	8.33e-11
22	140	22.9	132	4	016609	(APOARGC).	1.30e-10
23	138	22.6	334	6	046507	PLASMINOGEN (FRAGMENT)	3.13e-10
24	138	22.6	411	4	015844	UROKINASE-TYPE PLASMIN	3.13e-10
25	138	22.6	431	4	016618	UROKINASE PRECURSOR (E	3.13e-10
26	135	22.1	714	5	002001	NEUROSPECIFIC RECEPTOR	1.17e-09
27	135	22.1	724	5	096391	RECEPTOR TYROSINE KINASE	1.17e-09
28	128	20.9	202	13	090675	TISSUE-TYPE PLASMINOGEN	2.41e-08
29	127	20.8	944	11	092138	RO2.	3.70e-08
30	124	20.3	210	4	013494	HGF AGONIST/ANTAGONIST	1.32e-07
31	123	20.1	214	6	09XT70	UROKINASE-TYPE PLASMIN	2.02e-07
32	122	20.0	211	11	055027	HEPATOCTE GROWTH FACT	3.08e-07
33	119	19.5	145	6	028911	APOLIPROTEIN A (FRAG	1.08e-06
34	116	19.0	111	6	077688	PROTHROMBIN PRECURSOR	3.76e-06
35	116	19.0	685	5	024488	NEUROTHROMBIN RECEPTOR	3.76e-06
36	116	19.0	806	5	017576	KIN-8 PROTEIN.	3.76e-06
37	116	19.0	937	11	092139	ROR1	3.76e-06
38	116	19.0	943	4	001974	PROTEIN-TYROSINE KINASE	3.76e-06
39	115	18.8	937	4	001973	PROTEIN-TYROSINE KINASE	5.67e-06
40	111	18.2	263	4	000318	PUTATIVE PROTEIN.	2.90e-05
41	111	18.2	385	5	025101	SERINE PROTEINASE.	2.90e-05
42	93	15.2	389	6	097887	REISSNER'S FIBER GLYCO	3.16e-02
43	93	15.2	1637	6	09XS08	SCO-SPODIN (FRAGMENT)	3.16e-02
44	91	14.9	420	13	090504	THROMBIN.	6.81e-02
45	90	14.7	607	13	091001	THROMBIN.	6.81e-02

As hit 1, maybe

RESULT 1	PRELIMINARY:	PRT:	946 AA.
ID 007153	AC 007153		
DT 01-NOV-1996 (TREMBLrel. 01, Created)			
DI 01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE RECEPTOR TYROSINE KINASE.			
OS Torpedo californica (Pacific electric ray).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;			
OC Elasmobranchii; Rajiformes; Torpedinidae; Torpedo.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE-ELECTRIC ORGAN;			
RX MEDLINE: 93219391.			
RA JENNINGS C.G., DYER S.M., BURDEN S.J.;			
RT "Muscle-specific tyrosine-related receptor with a Kringle domain defines a			
RT distinct class of receptor tyrosine kinases."			
RL Proc. Natl. Acad. Sci. U.S.A. 90:2895-2899(1993).			
DR EMBL: L11311; AAA49285.1; .			
DR HSSP: P08631; 2HCK.			
DR PROSITE: PS00107; PROTEIN_KINASE_ATP. 1.			
DR PROSITE: PS00109; PROTEIN_KINASE_TYR. 1.			
DR PFAM: PF00047; 1g; 3.			
DR PFAM: PF00069; pkinase; 1.			
DR PFAM: PF00051; Kringle; 1.			
DR PRINTS: PR00018; KRINGLE.			
DR PRINTS: PR00109; TYRKINASE.			
SQ SEQUENCE 946 AA; 105892 MW; AF2A47E4 CRC32;			
Query Match	33.9%	Score 207;	DB 13; Length 946;
Best Local Similarity	39.4%	Pred. No. 3.09e-24;	
Matches 26; Conservative	13;	Mismatches 23;	Indels 4; Gaps 4;
DB 476 VNTVSGISGCRMSQAHFHRRLPEIPPEL-ANSDFCRNNGGSESPWC-YTMDRD-I 532			
OY 47 VSVTFGAPCLRMALVPEFLSPASNAQLGQNHNCPSDAG-RPMCFYGDARGKV 105			
DB 533 RWEFCN 538			
OY 106 DWGYCD 111			
RESULT 2	PRELIMINARY:	PRT:	597 AA.
ID 035727			

```

AC      035727. (Tremblrel. 05, Created)
DT      01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT      01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT      01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE      FACTOR XII.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RA      SCHLOESSER M., SCHWAGER S., ENGEL W.;
RL      Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL: X99571; CA67891.1; -.
DR      HSSP: P00760; 1A07.
DR      PROSITE: PS00023; FIBRONECTIN_2; 1.
DR      PROSITE: PS01253; FIBRONECTIN_1; 1.
DR      PFAM: PF00040; fn2; 1.
DR      PFAM: PF00089; trypsin; 1.
DR      PFAM: PF00051; kringle; 1.
DR      PFAM: PF00008; EGF; 2.
DR      PFAM: PF00039; fn1; 1.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
DR      PRINTS: PR00018; KRINGLE.
SQ      SEQUENCE 597 AA; 65638 MW; 1EA8CD44 CRC32;

Query Match          30.8%; Score 188; DB 11; Length 597;
Best Local Similarity 40.0%; Pred. No. 3,036-20;
Matches 26; Conservative 13; Mismatches 20; Indels 6; Gaps 6;

Db      230 GTTGCAGCCGATVATATRNMTENQALSWG-L-GH-HAFGCPNDPRNCFVWSG-DRL 285
Qy      48 SVTDEGACPLRAEVPFLERSP-PA-SWAQLRGRNHFRCSPDGAPWCYGDARGKY 105
Db      286 SMDYC 290
Qy      106 DWGTC 110

RESULT 3
ID      P70521 PRELIMINARY; PRT; 716 AA.
AC      P70521;
DT      01-FEB-1997 (Tremblrel. 02, Created)
DT      01-FEB-1997 (Tremblrel. 02, Last sequence update)
DE      01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE      HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR
DE      (MACROPHAGE STIMULATORY PROTEIN) (MSP).
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RA      OHSHIRO K., IWANA A., MATSUNO K., EZAKI T., SAKAMOTO O., HAMAGUCHI I.,
RA      KANASU N., SUDA T.;
RL      Molecular cloning of rat macrophage-stimulating protein and its
RL      involvement in the male reproductive system".
RL      Biochem. Biophys. Res. Commun. 227:773-280(1996)
CC      1- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
CC      CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
CC      CONSERVED.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
DR      EMBL: X95096; CA64473.1; -.
DR      HSSP: P00747; 1PK1.
DR      PFAM: PF00051; kringle; 4.
DR      PFAM: PF00089; trypsin; 1.
DR      PRINTS: PR00018; KRINGLE.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
KW      Signal.
TT      SIGNAL 1 31 POTENTIAL.

```

```

SQ CHAIN 32 716 MACROPHAGE STIMULATING PROTEIN.
ET SEQUENCE 716 AA; 80733 MW; 7C8133F6 CRC32;

Query Match 28.3%; Score 173; DB 11; Length 716;
Best Local Similarity 33.8%; Pred. No. 3,72e-17;
Matches 22; Conservative 16; Mismatches 24; Indels 3; Gaps 2

Db 203 VDVTSSGECRCRMDLQHSHSPHFPEKPPD-KALKDNYCRRPDASERPWCCTTDP--NVE 259
QY 47 VSVYDEGAPCLRMALVEVPFLERSPPASWAQLRGGRHNHCRSPDAGRWCFYGDARGKVD 106
Db 260 REFCD 264
QY 107 WCYCD 111

RESULT 4 PRELIMINARY; PRT; 560 AA.
ID 014520
AC 014520; 000653;
DT 01-NOV-1996 (TRENBLREL 01, Created)
DT 01-NOV-1996 (TRENBLREL 01, Last sequence update)
DT 01-NOV-1999 (TRENBLREL 12, Last annotation update)
DE HGF ACTIVATOR LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euteria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA RA
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX RX
RA MEDLINE; 96425001.
RA CHOI-MIURA N.H., TOBE T., SUMIYA J., NAKANO Y., SANO Y., MAZDA T.,
RA TOMITA M.;
RT "Purification and characterization of a novel hyaluronon-binding
RT protein (PHB) from human plasma. It has three EGF, a kringle and a
RT serine protease domain, similar to hepatocyte growth factor
RT activator.";
DR J. Biochem. 119:1157-1165(1996).
DR EMBL; D49742; BAA08576.1; -.
DR EMBL; S83182; AAB46909.1; -.
DR HSSP; P00763; IDPO.
DR HSSP; P00008; EGF; 3.
DR PFAM; PF00051; kringle; 1.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00018; KRINGLE.
DR PRINTS; PR00722; CHYMOTRYPSIN.
SO SEQUENCE 560 AA; 62671 MW; 4AC81907 CRC32;

Query Match 27.3%; Score 167; DB 4; Length 560;
Best Local Similarity 56.7%; Pred. No. 6,13e-16;
Matches 17; Conservative 4; Mismatches 9; Indels 0; Gaps 0.

Db 243 HNECRNDPDADEKPCWCFIKVINDKVKWEXCD 272
QY 82 HNECRSPDAGRPWCYFGDARGKVDWGYCD 111

RESULT 5 PRELIMINARY; PRT; 616 AA.
ID 097507
AC 097507;
DT 01-MAY-1999 (TRENBLREL 10, Created)
DT 01-MAY-1999 (TRENBLREL 10, Last sequence update)
DT 01-NOV-1999 (TRENBLREL 12, Last annotation update)
DE Fx11.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euteria; Cetartiodactyla; Suidae; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA RA
RL TISSUE-LIVER.
RA TAKAHASHI T., KIHARA T.;

```

RT "Porcine liver factor XII."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB022426; BAA37148.1; -
 DR HSSP: P00763; IDPO.
 DR PROSITE: PS00023; FIBRONECTIN_2; 1.
 DR PROSITE: PS01253; FIBRONECTIN_1; 1.
 SQ SEQUENCE 616 AA; 68012 MW; 31CCD856 CRC32;

Query Match 27.0%; Score 165; DB 6; Length 616;
 Best Local Similarity 40.0%; Pred. No. 1,55e-15;
 Matches 24; Conservative 11; Mismatches 19; Indels 6; Gaps 6;

DB 235 GACPQASATYWNMTAEOALWG-L-GD-HAFCRNPNDTRPCFWRG-DOLSWQYC 290
 53 GACPRLMAEVPFLERSP-PA-SWAQLRGQRHNFCRSPDGAGRWCYGDARGKVDWGYC 110

002935 PRELIMINARY; PRT; 290 AA.

AC 002935;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
 DE HEPATOCYTE GROWTH FACTOR, HEAVY CHAIN PRECURSOR.
 GN HGF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RA MEDLINE: 91200041.
 RA MIYAZAWA K., KITAMURA A., NAKA D., KITAMURA N.;
 RT "An alternatively processed mRNA generated from human hepatocyte
 growth factor gene";
 RL Eur. J. Biochem. 197;15-22(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93087571.
 RA HARTMANN G., NALDINI L., WEIDNER K.M., SACHS M., VIGNA E.,
 RA CONGOLLO P.M., BIRDMETER W.;
 RT "A functional domain in the heavy chain of scatter factor/hepatocyte
 growth factor binds the c-Met receptor and induces cell dissociation
 but not mitogenesis";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11574-11578(1992).
 CC -1- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
 HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPIC FACTOR, AND ACTS AS
 GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES. IT
 HAS NO DETECTABLE PROTEASE ACTIVITY.
 -1- SIMILARITY: CONTAINS TWO KRINGLE REGIONS.
 -1- SIMILARITY: HIGH, TO OTHER HGF, LOWER, TO PLASMINOGEN.
 DR EMBL: X57574; CAA40802.1; -
 DR EMBL: 102931; AAA52649.1; -
 DR HSSP: P14210; IBHT.
 DR PFAM: PF00051; kringie; 2.
 DR PRINTS: PR00018; KRINGLE.
 KW Growth factor; Kringie; Signal.
 KM Growth factor; Kringie; Signal.
 FT CHAIN 1 31
 FT SIGNAL 32 290
 FT DOMAIN 128 206
 FT DOMAIN 211 288
 FT MOD_RES 32 32
 FT PYRROLIDONE CARBOXYLIC ACID
 FT (BY SIMILARITY).
 SO SEQUENCE 290 AA; 33765 MW; 9EF113DE CRC32;

Query Match 26.7%; Score 163; DB 4; Length 290;
 Best Local Similarity 36.1%; Pred. No. 3,92e-15;
 Matches 22; Conservative 11; Mismatches 25; Indels 3; Gaps 2;

DB 226 TESGKICQRMWDQTPRHKFLPERYPD-KGFDNYCRNPDGQRPWCYTLDPHTR--WEY 282
 50 TDFGAPCLRMAEVPFLERSPPASMAQLRGQRHNFCRSPDGAGRWCYGDARGKVDWGY 109

DB 283 C 283
 QY 110 C 110

RESULT 7
 ID 014519 PRELIMINARY; PRT; 296 AA.

AC 014519;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
 DE COMPETITIVE HGF ANTAGONIST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CHAN A.M.L., ROBIN J.S., BOTTARO D.P., HIRSCHFELD D.W., CHEDID M.,
 RA AARONSON S.A.;
 RL Science 0:0-0(0).
 DR EMBL: M77227; AAA35980.1; -
 DR HSSP: P14210; IBHT.
 DR PFAM: PF00051; kringie; 2.
 DR PRINTS: PR00018; KRINGLE.
 SQ SEQUENCE 296 AA; 34546 MW; 202394A7 CRC32;

Query Match 26.7%; Score 163; DB 4; Length 296;
 Best Local Similarity 36.1%; Pred. No. 3,92e-15;
 Matches 22; Conservative 11; Mismatches 25; Indels 3; Gaps 2;

DB 226 TESGKICQRMWDQTPRHKFLPERYPD-KGFDNYCRNPDGQRPWCYTLDPHTR--WEY 282
 50 TDFGAPCLRMAEVPFLERSPPASMAQLRGQRHNFCRSPDGAGRWCYGDARGKVDWGY 109
 DB 283 C 283
 QY 110 C 110

RESULT 8
 ID 028398 PRELIMINARY; PRT; 2869 AA.

AC 028398;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
 DE APOLIPOPROTEIN (FRAGMENT).
 OS Eriaceus europaeus (Western European hedgehog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Insectivora; Erinaceidae; Erinaceinae; Erinaceus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 96025778.
 RA LAWN R.M., BOONMARK N.W., SCHWARTZ K., LINDAHL G.E., WADE D.P.,
 RA BYRNE C.D., FONG K.J., MEER K., PATTY L.;
 RT "The recurring evolution of lipoprotein(a). Insights from cloning of
 hedgehog apolipoprotein(a).";
 RL J. Biol. Chem. 270:24004-24009(1995).
 DR EMBL: U33170; AAC48522.1; -
 DR HSSP: P00747; IPMK.
 DR PFAM: PF00051; kringie; 31.
 KW Lipoprotein.
 FT NON_TER 1 1
 SO SEQUENCE 2869 AA; 318601 MW; FCC166B9 CRC32;

Query Match 26.5%; Score 162; DB 6; Length 2869;
 Best Local Similarity 35.9%; Pred. No. 6,22e-15;
 Matches 23; Conservative 9; Mismatches 29; Indels 3; Gaps 3;

DB 2417 MAYTASGHICQRMRESPPSHSHTPENYPT-KLVNVCNRPDGEAFAPNC-YTTNSA-VR 2473
 47 VSVTDGAPCLRMAEVPFLERSPPASMAQLRGQRHNFCRSPDGAGRWCYGDARGKVD 106

DB 2474 WEXC 2477
OY 107 WGYC 110

RESULT 9
ID 013308 PRELIMINARY: PRT: 567 AA.

AC 013208:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN HOMOLOG (Df151S1).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RT "Submitted (May 1995) to the EMBL/GenBank/DBJ databases.
MBL: U28054; AAC63092.1; -."
HSSP: P00747; 2PK4.
PFAM: PF00051; kringie; 4.
DR PRINTS: PR00018; KRINGLE.
SQ SEQUENCE 567 AA; 64116 MW; C2A95547 CRC32;

Query Match 26.4%; Score 161; DB 4; Length 567;
Best Local Similarity 46.7%; Pred. No. 9,86e-15;
Matches 21; Conservative 3; Mismatches 20; Indels 1; Gaps 1;

DB 276 GVPCCRMDAOIPIHQHFTPEKYAC-KDLRENFGRNPDGSAFPCF 319
OY 53 GACPLRMAEVPFLERSPPASMAQLRGQRHNFGRSPDGAAPWCF 97

RESULT 10
ID 014870 PRELIMINARY: PRT: 711 AA.

AC 014870:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE MACROPHAGE-STIMULATING PROTEIN PRECURSOR.
GN MST1.

OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER:
MEDLINE: 93340141.

OSHIKURA T., YUHKI N., WANG M.H., SKEEL A., LEONARD E.J.;
Cloning, sequencing, and expression of human macrophage stimulating
protein (MSP, MST1) confirms MSP gene as a member of the family of kringie
proteins and locates the MSP gene on chromosome 3.";
RL J. Biol. Chem. 268:15461-15468(1993).
DR EMBL: L11924; AA559872.1; -.
DR HSSP: P00747; 2PK4.
DR PFAM: PF00051; kringie; 4.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PR00018; KRINGLE.
DR PRINTS: PR00722; CHYMOTRYPSIN.

KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 483 POTENTIAL.
FT CHAIN 484 711 MACROPHAGE-STIMULATING PROTEIN.
FT CHAIN 711 MACROPHAGE-STIMULATING PROTEIN.
SQ SEQUENCE 711 AA; 80353 MW; 1445C010 CRC32;

Query Match 26.4%; Score 161; DB 4; Length 711;
Best Local Similarity 46.7%; Pred. No. 9,86e-15;
Matches 21; Conservative 3; Mismatches 20; Indels 1; Gaps 1;

DB 301 GVPCCRMDAOIPIHQHFTPEKYAC-KDLRENFGRNPDGSAFPCF 344
OY 53 GACPLRMAEVPFLERSPPASMAQLRGQRHNFGRSPDGAAPWCF 97

OY 53 GACPLRMAEVPFLERSPPASMAQLRGQRHNFGRSPDGAAPWCF 97

RESULT 11
ID 064007 PRELIMINARY: PRT: 728 AA.

AC 064007:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE HEPATOCYTE GROWTH FACTOR.
GN HGF.

OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94363381.
RA LEE C.C., KOZAK C.A., YAMADA K.M.;
RT "Structure, genetic mapping, and expression of the mouse Hgf/scatter
factor gene.";
RT Cell Adhes. Commun. 1:101-111(1993).
DR EMBL: S71816; AAB31855.1; -.
DR HSSP: P14210; 1BHT.
DR MGD: MGT196079; Hgf.
DR PFAM: PF00051; kringie; 4.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PR00018; KRINGLE.
DR PRINTS: PR00722; CHYMOTRYPSIN.

Query Match 26.4%; Score 161; DB 11; Length 728;
Best Local Similarity 36.1%; Pred. No. 9,86e-15;
Matches 22; Conservative 10; Mismatches 26; Indels 3; Gaps 3;

DB 227 TEGSKTCGRMDQOTPRHNFPERYPD-KGPDNCCRNPDGKRPBPCYLDP-DI-PMWY 283
OY 50 TDFGACPLRMAEVPFLERSPPASMAQLRGQRHNFGRSPDGAAPWCF 109

DB 284 C 284
OY 110 C 110

RESULT 12
ID 061662 PRELIMINARY: PRT: 748 AA.

AC 061662:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE HEPATOCYTE GROWTH FACTOR (FRAGMENT).
GN HGF.

OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6SJLF1/J; TISSUE=LIVER:
MEDLINE: 94060105.

RA LIU Y., MICHALOPOULOS G.K., ZARNEGAR R.;
RT "Molecular cloning and characterization of cDNA encoding mouse
hepatocyte growth factor.";
RT Biochim. Biophys. Acta 1216:299-303(1993).
DR EMBL: X72307; CAAS1054.1; -.
DR HSSP: P14210; 1BHT.
DR MGD: MGT196079; Hgf.
DR PFAM: PF00051; kringie; 4.
DR PFAM: PF00089; trypsin; 1.
FT NON_TER 1 1
SQ SEQUENCE 748 AA; 85200 MW; 24AE0820 CRC32;

Query Match 26.4%; Score 161; DB 11; Length 748;
Best Local Similarity 36.1%; Pred. No. 9,86e-15;
Matches 22; Conservative 10; Mismatches 26; Indels 3; Gaps 3;

Db 247 TSGKTCQRMDOQTPRHKFLPERYPD-KGFDDNYCRNPDGKRPWCYTLPD-DT-PMEX 303
QY 50 TDFGAPCLMAEVPFLERSPPASWAQLRGQRHNCRSPDAGRPWCYGDARGVDMGY 109
Db 304 C 304
QY 110 C 110

RESULT 13
ID 018783 PRELIMINARY; PRT: 806 AA.
AC 018783;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE LASMINGEN.
OC Macropus eugenii (Tammur wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Metatheria; Diprotodontia; Macropodidae; Macropus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER.

RA LAMN R.M., SCHWARTZ K., PATTHY L.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF012297; AAB65760.1;
DR HSSP; P00747; SHPG.
DR PFAM; PF00051; trypsin; 5.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PRO0018; KRINGLE.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
SQ SEQUENCE 806 AA; 90980 MW; 4371C1C6 CRC32;

Query Match 26.2%; Score 160; DB 6; Length 806;
Best Local Similarity 32.3%; Pred. No. 1,56e-14;

Matches 21; Conservative 13; Mismatches 28; Indels 3; Gaps 3;

Db 197 ISKTESGIEGCPWDSQEPHSHEIYPSKPS-KDLKENCYCRNPDGPRPWCFTSNP-EKR- 253
QY 47 VSVYDFGAPCLMAEVPFLERSPPASWAQLRGQRHNCRSPDAGRPWCYGDARGVDMGY 106
Db 254 WEFCN 258
QY 107 WGYCD 111

T 14
QY 1402 PRELIMINARY; PRT: 710 AA.
QY 1402;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HEPATOCYTE GROWTH FACTOR (FRAGMENT).
OC Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;
CC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95267690.
RA NAKAMURA H., TASHIRO K., NAKAMURA T., SHIOKAWA K.;
DE "Molecular cloning of Xenopus HGF cDNA and its expression studies in
Xenopus early embryogenesis."
DE EMBL; S77422; AAB34534.1;
DR HSSP; P14210; IBHT.
DR PFAM; PF00051; kringle; 4.
DR PFAM; PF00089; trypsin; 1.
DT NON TER 710
SQ SEQUENCE 710 AA; 81487 MW; 7911602F CRC32;

Query Match 26.0%; Score 159; DB 13; Length 710;
Best Local Similarity 34.4%; Pred. No. 2.47e-14;

Matches 21; Conservative 10; Mismatches 27; Indels 3; Gaps 3;
Db 213 TSGKTCQRMDOQTPRHKFLPERYPD-KGINDNYCRNPDGKRPWCYTLPD-DT-SMEF 269
QY 50 TDFGAPCLMAEVPFLERSPPASWAQLRGQRHNCRSPDAGRPWCYGDARGVDMGY 109
Db 270 C 270
QY 110 C 110

RESULT 15
ID 046506 PRELIMINARY; PRT: 454 AA.
AC 046506;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE APOLIPOPROTEIN A (FRAGMENT).
GN BABAPOL.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Papio.
RN [1]
RP SEQUENCE FROM N.A.
RA COX L.A., JETT C., HIXSON J.E.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF029691; AAB97886.1;
DR HSSP; P00747; 2PK4.
DR PFAM; PF00051; kringle; 2.
DR PFAM; PF00089; trypsin; 1.
KW Lipoprotein.
FT NON TER 1
SQ SEQUENCE 454 AA; 50041 MW; 2400BE6C CRC32;

Query Match 24.9%; Score 152; DB 6; Length 454;
Best Local Similarity 33.9%; Pred. No. 6.02e-13;

Matches 20; Conservative 12; Mismatches 24; Indels 3; Gaps 3;

Db 6 GRNCQASMTPHQSRTPKXNPNA-GLTRNYCRNPDGPRPWCYTLPD-S-YRREYCN 61
QY 53 GAPCLMAEVPFLERSPPASWAQLRGQRHNCRSPDAGRPWCYGDARGVDMGY 111

Search completed: Mon Mar 13 10:21:47 2000
Job time : 17 secs.

THIS PAGE BLANK (USPTO)

potenz

Cherch Sidiyasa Sarbut 2

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h-pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:19:31 2000; MasPar time 8.11 Seconds
 391,512 Million cell updates/sec
 Tabular output not generated.

```

Title:          >US-09-147-947-6
Description:    (578-822) from US09147947A.ppt (1 of 6)
Perfect Score: 1892
Sequence:      1 IIGKNSLRCGMPQVSLRI.....PGVYIKVSAFPMWIKSVTKL 24F

```

Scoring table: PAM 150

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

```
Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1
```

Statistics: Mean 30.953; Variance 131.070; scale 0.236

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

ID	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	594	31.4	638	2	US-08-681-	Sequence 3, Applicatio	4.656-46
2	554	29.3	356	2	US-08-681-	Sequence 1, Applicatio	4.456-46
3	545	28.8	400	1	US-07-942-	Sequence 3, Applicatio	4.726-46
4	543	28.7	306	2	US-08-560-	Sequence 45, Applicati	4.326-46
5	543	28.7	331	2	US-08-560-	Sequence 46, Applicati	4.332-46
6	543	28.7	365	1	US-08-120-	Sequence 83, Applicati	4.332-46
7	543	28.7	365	1	US-08-093-	Sequence 83, Applicati	4.326-46
8	543	28.7	333	2	US-08-060-	Sequence 44, Applicatio	4.326-46
9	543	28.7	411	1	US-08-087-	Sequence 1, Applicatio	4.326-46
10	543	28.7	411	1	US-08-086-	Sequence 18, Applicati	4.326-46
11	543	28.7	411	1	US-08-153-	Sequence 18, Applicati	4.326-46
12	543	28.7	400	4	52195656-1	Patent No. 5219569.	4.332-46
13	543	28.7	431	4	5188829-1	Patent No. 5188829.	4.326-46
14	543	28.7	432	2	US-08-560-	Sequence 47, Applicati	4.326-46
15	534	28.2	367	2	US-08-811-	Sequence 1, Applicatio	4.086-40
16	533	28.2	335	2	US-08-811-	Sequence 47, Applicati	4.086-40
17	533	28.2	437	2	US-08-811-	Sequence 51, Applicati	4.086-40
18	533	28.2	437	2	US-08-811-	Sequence 51, Applicati	4.086-40
19	533	28.2	537	2	US-08-811-	Sequence 39, Applicati	8.006-40
20	530	28.0	354	2	US-08-560-	Sequence 48, Applicati	8.006-40
21	530	28.0	411	2	US-08-560-	Sequence 48, Applicati	1.576-33
22	527	27.9	354	2	US-08-017-	Sequence 61, Applicati	1.576-33

24	527	27.9	355	4	5223556-1	Patent No. 5223556	1.57e-33
25	527	27.9	355	1	US-08-217-	Sequence 1, Applicatio	1.57e-33
26	527	27.9	355	1	US-08-47-	Sequence 6, Applicatio	1.57e-33
27	527	27.9	355	2	US-08-41-	Sequence 45, Applicati	1.57e-33
28	527	27.9	355	1	US-08-17-	Sequence 2, Applicatio	1.57e-33
29	527	27.9	355	1	US-08-217-	Sequence 1, Applicatio	1.57e-33
30	527	27.9	355	1	US-08-47-	Sequence 8, Applicatio	1.57e-33
31	527	27.9	356	1	US-08-47-	Sequence 4, Applicatio	1.57e-33
32	527	27.9	356	1	US-08-58-	Sequence 6, Applicatio	1.57e-33
33	527	27.9	389	2	US-08-81-1-	Sequence 67, Applicati	1.57e-33
34	527	27.9	437	2	US-08-81-1-	Sequence 55, Applicati	1.57e-33
35	527	27.9	527	1	5185259-8	Patent No. 518559	1.57e-33
36	527	27.9	527	1	US-07-609-	Sequence 16, Applicati	1.57e-33
37	527	27.9	527	3	PCT-US91-0	Sequence 2, Applicatio	1.57e-33
38	527	27.9	527	3	5244773-2	Patent No. 5244773	1.57e-33
39	527	27.9	562	4	5244676-5	Patent No. 5244676	1.57e-33
40	527	27.9	562	2	US-08-563-	Sequence 50, Applicati	1.57e-33
41	527	27.9	562	2	US-08-880-	Sequence 38, Applicati	1.57e-33
42	527	27.9	562	4	5185259-3	Patent No. 518559	1.57e-33
43	527	27.9	562	4	5200340-2	Patent No. 5200340	1.57e-33
44	527	27.9	562	2	US-08-81-1-	Sequence 43, Applicati	1.57e-33
45	527	27.9	562	2			

RESULT 1
ID US-08-681-151-3 STANDARD; PRT; 638 AA..
XX
XX
XXXXXX
DT
XT
Sequence 3, Application US/08681151
DE
XX
Sequence 3, Application US/08681151
CC Patent No. 5869637
CC GENERAL INFORMATION:
CC APPLICANT: Au-Young, Janice
CC APPLICANT: Bandman, Olga
CC APPLICANT: Braxton, Scott Michael
CC APPLICANT: Goli, Surya
CC TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
CC STREET: 3174 Porter Drive
CC City: Palo Alto
CC STATE: CA
CC COUNTRY: US
CC COUNT: 1
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: diskette
CC COMPUTER: IBM compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ Version 1.5
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/681,151
CC FILING DATE: Herewith
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0074US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:

CC APPLICANT: HEINZEL-WIELAND, Regina
CC APPLICANT: STEFFENS, Gerd Josef
CC TITLE OF INVENTION: Proteins having fibrinolytic and
CC TITLE OF INVENTION: Coagulation-inhibiting Properties
CC NUMBER OF SEQUENCES: 60
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
CC STREET: 1200 G Street, N.W., Suite 700
CC CITY: Washington
CC STATE: DC
CC COUNTRY: USA
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30 (FPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/560, 098A
CC FILING DATE: 17-NOV-1995
CC PRIORITY APPLICATION DATA:
CC APPLICATION NUMBER: P 44 40 892.7
CC FILING DATE: 17-NOV-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: EVANS, Joseph D.
CC REGISTRATION NUMBER: 26,269
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 628-8800
CC TELEFAX: (202) 628-8844
CC INFORMATION FOR SEQ ID NO: 46:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 331 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 331 AA: 37442 MW; 581384 CN;
SQ
Query Match 28.7%; Score 543; DB 2; Length 331;
Best Local Similarity 36.9%; Pred. No. 4.32e-41;
Matches 93; Conservative 60; Mismatches 85; Indels 14; Gaps 13;
Db 54 IIGGEFTTENQWFAAIRRHGSGVTVYVCGSLISPCWVISATHCFTIDYPRKE-DYIV 112
113 IIGGKNSLRGMPQVSLRLKSSHGDRLLCGATLLSSCWLTAAHCFKRGSTSYAV 637
578 IIGGKNSLRGMPQVSLRLKSSHGDRLLCGATLLSSCWLTAAHCFKRGSTSYAV 637
113 YLGRSLNSTOGEMKEVENLILHKDYSADTLAHNDIALILKRSKGCQAOPSTIQT 172
638 RVGDYHFLVPEEFEEELIGVOOIVIHREYRPR-SDY-DIALVRLQGPEDCCARFSSHVLP 695
173 ICLPSMYND-PQ-FGTSCETIGFGKENSVDLYPEOLKMTVVKLISHRECOOPHYGSEV 230
696 ACLP-LMREPOKTASNYITGNG-DTGRAY--SRLQQAAILPLKRFCEE-RYGR-F 749
231 TTKMLCAAD-PQMK-TDSCGDSGGPLVCSLOGRM-TLTGIVSWGRCALDKRPGYTRY 287
750 TGMRLCAGNLHHRKRVDSGCGPLMCRPESWVYGVTSWGYCGVXDSPGYTRY 809
Db 288 SHPLPWIRSHTK 299
QY 810 SAFVPMIKSVTK 821
RESULT 6
ID US-08-720-012-83 STANDARD: PRT: 365 AA.
AC xxxxxx
DT
DE Sequence 83, Application US/08720012
XX

CC Sequence 83, Application US/08720012
CC Patent No. 574/291
CC GENERAL INFORMATION:
CC APPLICANT: STEFFENS, Gerd J.
CC APPLICANT: WENDEL, STEPHAN
CC APPLICANT: SCHNEIDER, JOHANNES
CC APPLICANT: HEINZEL-WIELAND, REGINA
CC APPLICANT: SAUNDERS, DEREK J.
CC TITLE OF INVENTION: BIFUNCTIONAL UROKINASE VARIANTS WITH
CC TITLE OF INVENTION: IMPROVED FIBRINOLYTIC CHARACTERISTICS AND THROMBIN
CC TITLE OF INVENTION: INHIBITING EFFECT
CC NUMBER OF SEQUENCES: 83
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
CC STREET: 1200 G Street, N. W. Suite 700
CC CITY: Washington, D.C.
CC COUNTRY: U. S.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/720, 012
CC FILING DATE: 27-SEP-1996
CC CLASSIFICATION: 435
CC PRIORITY APPLICATION DATA:
CC APPLICATION NUMBER: US 08/093,741
CC FILING DATE: 20-JUL-1993
CC APPLICATION NUMBER: DE P43 23 754.1
CC FILING DATE: 15-JUL-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: EVANS, Joseph D.
CC REGISTRATION NUMBER: 26,269
CC REFERENCE/DOCKET NUMBER: 148/41345
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202)628-8800
CC TELEFAX: (202)628-8844
CC INFORMATION FOR SEQ ID NO: 83:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 365 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC SEQUENCE 365 AA: 41229 MW; 706998 CN;
SQ
Query Match 28.7%; Score 543; DB 1; Length 365;
Best Local Similarity 36.9%; Pred. No. 4.32e-41;
Matches 93; Conservative 60; Mismatches 85; Indels 14; Gaps 13;
Db 113 IIGGEFTTENQWFAAIRRHGSGVTVYVCGSLISPCWVISATHCFTIDYPRKE-DYIV 171
578 IIGGKNSLRGMPQVSLRLKSSHGDRLLCGATLLSSCWLTAAHCFKRGSTSYAV 637
172 YLGRSLNSTOGEMKEVENLILHKDYSADTLAHNDIALILKRSKGCQAOPSTIQT 231
638 RVGDYHFLVPEEFEEELIGVOOIVIHREYRPR-SDY-DIALVRLQGPEDCCARFSSHVLP 695
232 ICLPSMYND-PQ-FGTSCETIGFGKENSVDLYPEOLKMTVVKLISHRECOOPHYGSEV 289
696 ACLP-LMREPOKTASNYITGNG-DTGRAY--SRLQQAAILPLKRFCEE-RYGR-F 749
290 TTKMLCAAD-PQMK-TDSCGDSGGPLVCSLOGRM-TLTGIVSWGRCALDKRPGYTRY 346
750 TGMRLCAGNLHHRKRVDSGCGPLMCRPESWVYGVTSWGYCGVXDSPGYTRY 809
Db 347 SHPLPWIRSHTK 358
QY 810 SAFVPMIKSVTK 821
RESULT 7
ID US-08-093-741-83 STANDARD: PRT: 365 AA.
AC
DT
DE
XX

CC ADDRESSEE: P.C.
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC CITY: ARLINGTON
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/811,949
CC FILING DATE: 05-MAR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OBLON, NORMAN F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 18-966-0
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-413-3000
CC TELEFAX: 703-413-2220
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 347 AA; 38594 MW; 612249 CN;

Query Match: 28.2%; Score 534; DB 2; Length 347;
Best Local Similarity 39.6%; Pred. No. 3,26e-40;
Matches 95; Conservative 52; Mismatches 76; Indels 17; Gaps 13;
DB 108 PMQALFAKHRSFGERFLCGILISSCWILSAHCFQERF--PPHLLVILGRTYR-VV 164
QY 590 PMQVSLRLKSSHG-DG-RLICGATILSSCWLTFAHCFK-RYGNSTRSYAVRVG-DYHTLV 646
DB 165 PEEFQKFEVEKYIVKFEFDDDTYNDIALIOLKSDSSRCQAESSVRYVCULP-ADLQL 223
QY 647 PEEFEEIGVQIVIHREYRPDRSDYDIALVRLQGEPEOCARFSSHVLPACLPWMREPO 706
DB 224 PMTECELSGYGKHEALSPYSERLKEAHVRLYPSSRCTSOHLINRTVTDNMLCAGDTRS 283
QY 707 KTASNCYITGWS--DT-GRAYSRITLQAAIPLPKRFC--ERYKGR-FTGRMLCAGNLHE 761
DB 284 GGPQANLHDACGGSGPLVCLNDGRMTLV-GIISWLGCGGQKDPGVYTKYTNVLDWIR 342
QY 762 ---HKRV-DSCGDSGGPLMCCERPGESEWVYGVTSWGCYCGKVDSPGVYTKVASAFVPMIK 817

Search completed: Mon Mar 13 10:19:40 2000
Job time : 9 secs.

THIS PAGE BLANK (USPTO)

FT DOMAIN 500 601 SRCR 4.
 FT DOMAIN 619 875 SERINE PROTEASE.
 FT DOMAIN 619 630 ZMOGEN ACTIVATION REGION.
 FT ACT SITE 630 631 REACTIVE BOND (POTENTIAL).
 FT ACT SITE 630 631 CHARGE RELAY SYSTEM.
 FT ACT SITE 726 726 CHARGE RELAY SYSTEM.
 FT ACT SITE 825 825 CHARGE RELAY SYSTEM.
 FT DISULFID 619 750 POTENTIAL.
 FT CARBOHYD 26 26 POTENTIAL.
 FT CARBOHYD 683 683 POTENTIAL.
 SQ SEQUENCE 875 AA: 97011 MW: 6752272B CRC32:

Query Match 100.0%; Score 1892; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 631 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSYAV 690
 578 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSYAV 637
 691 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRDSYDIALVRLQGEBOCARFSSHYLPAC 750
 638 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRDSYDIALVRLQGEBOCARFSSHYLPAC 697
 Db 751 LPLMRERPKTASNCYITGMDTGRAVSRLLQQAIPLPKRCCEERYKGRFTGMCLAG 810
 698 LPLMRERPKTASNCYITGMDTGRAVSRLLQQAIPLPKRCCEERYKGRFTGMCLAG 757
 Qy 811 NLHEHRRVDSGCGDGGPMLCERPGESWVYGVTSWGYGCGVDSPPVYTKVSAFVPMIK 870
 758 NLHEHRRVDSGCGDGGPMLCERPGESWVYGVTSWGYGCGVDSPPVYTKVSAFVPMIK 817
 Qy 871 SVTKL 875
 818 SVTKL 822

RESULT 2
 ID NEUR_MOUSE STANDARD: PRT: 761 AA.
 AC 008762;
 DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE NEURORXPSIN PRECURSOR (EC 3.4.21.-) (MOTOPISIN) (BRAIN-SPECIFIC SERINE
 PROTEASE 3) (BSP-3).
 DE PRS12.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 [1] SEQUENCE FROM N.A.
 [1] TISSUE-BRAIN;
 MEDLINE: 97401523.
 RA GICHENMEND T.P., KUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
 RT "Neurocrystin, a novel multidomain serine protease expressed in the
 RT nervous system";
 RT Mol. Cell. Neurosci. 9:207-219(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98008848.
 RA YAMAMURA Y., YAMASHIRO K., TSUBOKA N., NAKAZATO H., TSUJIMURA A.,
 RA YAMAGUCHI N.;
 RT "Molecular cloning of a novel brain-specific serine protease with a
 RT kringle-like structure and three scavenger receptor cysteine-rich
 RT motifs";
 RT Biochem. Biophys. Res. Commun. 239:386-392(1997).
 CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
 CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
 CC LEARNING AND MEMORY OPERATIONS.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
 CC AND AMYGALA.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPsin FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y13192; CAA3646.1;
 CC EMBL: D89871; BAA23986.1;
 CC MGD: MGI:1100881; PRSS12.
 CC DR PFAM: PF00530; SRCR; 3.
 CC DR PFAM: PF00089; trypsin; 1.
 CC DR PROSITE: PS00134; TRYPsin_HIS; 1.
 CC DR PROSITE: PS00135; TRYPsin_SER; 1.
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.
 CC KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; signal.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT DOMAIN 22 761
 FT ACT_SITE 85 157
 FT ACT_SITE 166 267
 FT ACT_SITE 273 373
 FT ACT_SITE 386 487
 FT ACT_SITE 505 761
 FT ACT_SITE 516 517
 FT ACT_SITE 562 562
 FT ACT_SITE 612 612
 FT ACT_SITE 711 711
 FT DISULFID 505 636
 FT CARBOHYD 93 93
 FT CARBOHYD 521 521
 FT CARBOHYD 569 569
 SQ SEQUENCE 761 AA: 3F3CAF35 CRC32:

Query Match 92.6%; Score 1752; DB 1; Length 761;
 Best Local Similarity 89.8%; Pred. No. 0.00e+00;
 Matches 220; Conservative 17; Mismatches 8; Indels 0; Gaps 0;

Db 517 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSYAV 576
 578 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSYAV 637
 Qy 637 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRDSYDIALVRLQGEBOCARFSSHYLPAC 696
 638 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRDSYDIALVRLQGEBOCARFSSHYLPAC 697
 Db 697 LPLMRERPKTASNCYITGMDTGRAVSRLLQQAIPLPKRCCEERYKGRFTGMCLAG 757
 698 LPLMRERPKTASNCYITGMDTGRAVSRLLQQAIPLPKRCCEERYKGRFTGMCLAG 757
 Qy 758 NLHEHRRVDSGCGDGGPMLCERPGESWVYGVTSWGYGCGVDSPPVYTKVSAFVPMIK 817
 757 SVTKL 761
 818 SVTKL 822

RESULT 3
 ID FALL_HUMAN STANDARD: PRT: 625 AA.
 AC P03951;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN
 DE ANTECEDENT) (PTA).
 GN F11.

QY 753 MLCAGNLEHHRVDSGQSGPLMCEPESWVYGVTSWGGCVKDSPGVYTVASAF 812
 Db 615 VDWI 618
 QY 813 VPWI 816
 RESULT 4
 ID KAL_MOUSE STANDARD: PRT: 638 AA.
 AC P26262:
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KIK3 OR PK.
 OS Mus musculus (Mouse).
 RA Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 RA Theria; Rodentia; Sciurognathi; Muridae; Mus.
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX STRAIN-BALB/C; TISSUE-LIVER:
 RX MEDLINE: 91090844.
 RA SEDDAH N.G., SAWYER N., HAMELIN J., MION P., BEAUBIEN G.,
 RA BRACHDAPA L., ROCHERONT J., MBIKAY M., CHRETIEN M.;
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species.";
 RL DNA Cell Biol. 9:737-748(1990).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HWM KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HWM KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC MBL: M58588; AAA6393.1; -
 CC LIR: A36557; KOMSPL.
 CC HSSP: P00750; LRPE.
 DR MGD: MGI:102849; KIK3.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00495; APPLE; 4.
 DR PFAM; PF00024; apple; 4.
 DR PFAM; PF00089; trypsin; 1.
 DR Hydrolase; Serine protease; Glycoprotein; Plasma; zymogen; Signal;
 KM Fibrolysis; Blood coagulation; Inflammatory response; Liver;
 KM Duplication; Bradykinin.
 FT SIGNAL 1 19
 FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
 FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
 FT REPEAT 20 105 APPLE 1.
 FT REPEAT 110 195 APPLE 2.
 FT REPEAT 200 285 APPLE 3.
 FT REPEAT 291 376 APPLE 4.
 FT DOMAIN 389 621 CATALYTIC.
 FT CARBOHYD 127 127 PROBABLE.
 FT CARBOHYD 308 308 PROBABLE.
 FT CARBOHYD 396 396 PROBABLE.

FT CARBOHYD 453 453 PROBABLE.
 FT CARBOHYD 494 494 PROBABLE.
 FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
 FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
 FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
 FT DISULFID 21 104 BY SIMILARITY.
 FT DISULFID 47 77 BY SIMILARITY.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 111 194 BY SIMILARITY.
 FT DISULFID 137 166 BY SIMILARITY.
 FT DISULFID 141 147 BY SIMILARITY.
 FT DISULFID 201 284 BY SIMILARITY.
 FT DISULFID 227 256 BY SIMILARITY.
 FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 318 347 BY SIMILARITY.
 FT DISULFID 322 328 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 363 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA; 71368 MW; 8370CD2E CRC32;
 Query Match 31.7%; Score 600; DB 1; Length 638;
 Best Local Similarity 40.6%; Pred. No. 3,41e-125;
 Matches 99; Conservative 54; Mismatches 73; Indels 18; Gaps 16;
 Db 391 YGCTASAGEMPWQVSLQVLY-SQTHL-CGSGITIGROWTITAHACDGIPIPYWRIY 448
 QY 578 IIGKNSLGGMPQVSLKSSHDGRLCGATLLSCWVITAAHCKR--RYGNSTRSY 635
 Db 449 GGLIS-L-SEIKKEPPSSR-IRELLIHQEKYSEGNVIALIKITPLN-YTEF--QK-P 501
 QY 636 AVAVGSHYLVBEFEELGVQOYIVHREYRDRBDYDALVRLQGPPEGCCARFESHLP 695
 Db 502 ICIP-SKADNTTYTNCVYWGVTKEGETONILQKATIPLVNPECQKRYDYINKQ 560
 QY 696 ACILPRLRPPQRTASNCYITGMDI-GRAYSR-T-LOQAIPPLPFRCEERYKGRFTGR- 752
 Db 561 MLCAGY-KEGG-TDACKGSGPLVCKHSG-RWLVGITSNMGEGGRKDOGVYTVASEY 617
 QY 753 MLCAGNLEHHRVDSGQSGPLMCEPESWVYGVTSWGGCVKDSPGVYTVASAF 812
 Db 618 MDMI 621
 QY 813 VPWI 816
 RESULT 5
 ID KAL_HUMAN STANDARD: PRT: 638 AA.
 AC P03952:
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KIK3.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86243359.
 RA CHUNG D.W., FUJIKAWA K., MCMULLEN B.A., DAVIE E.W.;
 RT "Human plasma prekallikrein, a zymogen to a serine protease that
 RT contains four tandem repeats.";
 RL Biochemistry 25:2410-2417(1986).
 RN [2]
 RP PARTIAL SEQUENCE. AND DISULFIDE BONDS.
 RX MEDLINE: 91152016.
 RA MCMULLEN B.A., FUJIKAWA K., DAVIE E.W.;

[illegible]

ID ACRO.RABIT STANDARD: PRT: 431 AA.
AC P48038;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1991 (Rel. 35, Last annotation update)
DE ACROSIN PRECURSOR (EC 3.4.21.10).
GN ACR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE; TISSUE-TESTIS;
RX MEDLINE: 94368861.
RA RICHARDSON R.T., O'RAND M.G.;
"Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel preproacrosin-related cDNA."
Biochim. Biophys. Acta 1219:215-218(1994).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY. IT
IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U05204; AAA61630.1;
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PFAM: PF00089; trypsin; 1.
KM Hydrolyase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16
FT CHAIN 17 431
FT CHAIN 17 39
FT CHAIN 17 39
FT CHAIN 40 7
PROPEP ? 431
DISULFID 22 160
DISULFID 26 160
DISULFID 71 87
DISULFID 175 244
DISULFID 207 223
DISULFID 234 264
CARBOHYD 19 19
CARBOHYD 208 208
ACT_SITE 86 86
ACT_SITE 140 140
ACT_SITE 238 238
FT ACT_SITE 238
FT ACT_SITE 238
SO SEQUENCE 431 AA; 46422 MW; D350A173 CRC32;
Query Match 29.38; Score 555; DB 1; Length 431;
Best Local Similarity 39.18; Pred. No. 3.97e-113;
Matches 99; Conservative 54; Mismatches 81; Indels 19; Gaps 13;

QY 692 HYPAFLPLMRERPOXTASCIYITGNG---DTGRATSRILQQAATPLLRKFE--ERIK 746
DB 215 GRITASNLGAGYPSG-K-IDTCGDSGGLMCRENGEPPVVOGITSWGCARARPGI 272
QY 747 GRITGMLCAGNLNHEKRPDSCGDSGGLPMC-ERRGESWYVGYVSMWCGGVKNSPCV 805
DB 273 YRTAFPLDMIAS 285
QY 806 YTKSAFVFWIKS 818
RESULT 11
ID ACRO.MOUSE STANDARD: PRT: 436 AA.
AC P23578;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACROSIN PRECURSOR (EC 3.4.21.10).
GN ACR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91185335.
RA KASHIWABARA S., BABA T., TAKADA M., WATANABE K., YANO Y., ARAI Y.;
"Primary structure of mouse proacrosin deduced from the cDNA sequence
RT and its gene expression during spermatogenesis."
J. Biochem. 108:785-791(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92041732.
RA WATANABE K., BABA T., KASHIWABARA S., OKAMOTO A., ARAI Y.;
"Structure and organization of the mouse acrosin gene."
J. Biochem. 109:828-833(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90255839.
RA KLEMM U., MAIER W.M., TSAOUSIDOU S., ADHAM I., WILLISON K.,
RA ENGEL W.;
"Mouse preproacrosin: cDNA sequence, primary structure and
RT postmeiotic expression in spermatogenesis."
J. Differentiation 42:160-166(1990).
RN [4]
RX SEQUENCE FROM N.A.
RA MEDLINE: 92147126.
RA KREMLING H., KEIME S., WILHELM K., ADHAM I.M., HAMFISTER H.,
RA ENGEL W.;
"Mouse proacrosin gene: nucleotide sequence, diploid expression, and
RT chromosomal localization."
Genomics 11:828-834(1991).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY. IT
IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: S66245; AAB20293.1;
DR EMBL: S64500; AAB20293.1; JOINED.

RESULT 2
ID Q9Y1V3 PRELIMINARY; PRT; 868 AA.
AC Q9Y1V3;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
GN TUNICATE RETINOIC ACID-INDUCIBLE MODULAR PROTEASE PRECURSOR.
OS Polyandrocampa misakiensis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Styelidae; Polyandrocampa.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE SPOT.
RA OHASHI M., KAMAMURA K., FUJII N., YUBISUI T., FUJIMARA S.;
"A retinoic acid-inducible modular protease in budding ascidians";
EMBL: AB030007; BAA82522.1; -
PROSITE: PS01209; LDRA.1; 3.
FT SIGNAL 1 28 POTENTIAL.
SQ SEQUENCE 868 AA; 97660 MW; 68282AE6 CRC32;
Query Match 32.8%; Score 620; DB 5; Length 868;
Best Local Similarity 39.5%; Pred. No. 2,25e-126;
Matches 100; Conservative 56; Mismatches 77; Indels 20; Gaps 13;
Db 624 IYGGSGEPHEPMQAGIMLPMWY---W--CGGSLIHPCWLTAAHCFVR-EYIRDYTI 677
QY 578 IIGKNSLRGMPQVSLRKSSHGDRLLCGATLLSSCWLTAAHCKRYGNSYRSTYAV 637
Db 678 RLGDHITGVDETEQLFKIAEIKH-DYVNTTKENDIALRIENDARCATITEVOYVC 736
QY 638 RVGDYHLYVEEFEEELGVQOIVIHREYRPRSDYDALVALQPEDECAFSSHVLPAC 697
Db 737 LP--KSSQFDATKICVGTGKDSATAVRAYVYLOEAEIPLIANKKCLRDSEYT-OLG 793
QY 698 LPLMRERPQ-KTASNCYITGWG-D-TG-RAYSRTLOQAAIPLPKRFE--EERYKGRFT 750
Db 794 PTFECAGYTLGK--DSCGDSGGPLSCROOSDRYVYVWGIVSGNGCAKRGYVAKV 851
QY 751 GRMLCAGNLHHRVDSGDSGGPLMC-ERPGESWVYVGTSMGCGVXDSPGVYTKV 809
Db 852 AVFIDWIEQMTGL 864
QY 810 SAFVPMIKSVTKL 822
Query Match 3
Q9XZM7; PRELIMINARY; PRT; 581 AA.
AC Q9XZM7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
GN CORTICAL GRANULE SERINE PROTEASE 1 PRECURSOR.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidae;
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RC HALEY S.A., WESSLE G.M.;
"The cortical granule serine protease CGSP1 of the sea urchin,
RT Strongylocentrotus purpuratus, is autocatalytic and has an LDL
receptor-like domain";
EMBL: AF149789; AAD37426.1; -
DEV. Biol. 0:0-0(1999).
DR PROSITE: PS01209; LDRA.1; 3.
KW Signal; Protease; Glycoprotein.
FT SIGNAL 1 29 POTENTIAL.
CHAIN 30 581 CORTICAL GRANULE SERINE PROTEASE 1.

SQ SEQUENCE 581 AA; 64438 MW; 3CA00786 CRC32;
Query Match 31.1%; Score 588; DB 5; Length 581;
Best Local Similarity 40.0%; Pred. No. 4.81e-118;
Matches 98; Conservative 55; Mismatches 78; Indels 14; Gaps 12;
Db 334 IYGGPATAQDMPQAOLEFRT-RGSQQLVGGTLLDPQVYVLTAAFCMGPMMATSRMOY 392
QY 578 IIGKNSLRGMPQVSLRKSSHGDRLLCGATLLSSCWLTAAHCKRYGNSYRSTYAV 637
Db 393 HLGHSDVDFPEAGSOHRLREIFVHKRFGHGGVCDIALLLIDPPVO-E--TGOIMW 449
QY 638 RVGDYHLYVEEFEEELGVQOIVIHREYRPRDS-DYDALVALQPEDECAFSSHVLP 695
Db 450 ACID--EGMPLNDETECYISGVTENGNGPDLHARRPLIPRRCYKNSYKIEK 507
QY 696 ACLPLMRERPQKTSNICYITGWSDTGRA-YSR-TLOQAAIPLPKRFE--ER-YKGRFTG 751
Db 508 TMLCAGL-EGG-IDACGDSGGPLSCGPDHMYVYVGTSMWGHCAIANKPGVYTKVS 565
QY 752 RMLCAGNLHHRVDSGDSGGPLMCERPGESWVYVGTSMGCGVXDSPGVYTKVSA 811
Db 566 YLDWI 570
QY 812 FVPMI 816
RESULT 4
ID Q29015 PRELIMINARY; PRT; 415 AA.
AC Q29015;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
GN PREPROACROSIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER.
RX MEDLINE: 96348713.
RA ADHAM I.M., KREMLING H., NIETER S., ZIMMERMANN S., HUMMEL M.,
SCHROETER U., ENGEL W.;
"The structures of the bovine and porcine proacrosin genes and their
RT conservation among mammals";
RL Biol. Chem. Hoppe-Seyler 377:261-265(1996).
DR EMBL: X58549; CAA41440.1; -
PRAM: PF00089; trypsin.1.
DR PRINTS: PR00722; CHYNOTRYPsin.
FT CHAIN 17 415 ACROSIN.
SQ SEQUENCE 415 AA; 45361 MW; 80A12323 CRC32;
Query Match 30.3%; Score 574; DB 6; Length 415;
Best Local Similarity 38.1%; Pred. No. 2.07e-114;
Matches 99; Conservative 56; Mismatches 79; Indels 19; Gaps 14;
Db 40 VVGMSAEPGAPMWSLIQTFMYHNNRHYTCGILLNSHWVLTAAHCKRKKYTDRL 99
QY 578 IIGKNSLRGMPQVSLRKSSHGDRLLCGATLLSSCWLTAAHCKRYGNSYR- 633
Db 100 IFGANEVYWSNKPVPLOERF-VEEIIHKEYVSGLEINDIALIKI-TPVPCGP-- 155
QY 634 SYAVR-VG-DYHTLVPEEFEEELGVQOIVIHREYRPRSDYDALVALQPEDECAFSS 691
Db 156 -IGGCLPQKAPRPPAPQVCTGWCYLKREKPRSPLOEARVALIDLELCSNRWYN 214
QY 692 HVLPACPLMRERPQKTSNICYITGWG--DTGRAYSRLOQAAIPLPKRFE--R-YK 746
Db 215 GRISTVWAGYPRG-K-IDTCGDSGGPLMCDRAENFTVVVGTSMWGHCAIANKPGVY 272
QY 747 GRITGRMLCAGNLHHRVDSGDSGGPLMC-ERPGESWVYVGTSMGCGVXDSPGV 805
Db 273 YTSWPLNMIAS 285

QY 806 YTKVSAFVPMIKS 818

RESULT 5 PRELIMINARY: PRT: 377 AA.

AC P79343;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-NOV-1997 (TREMblrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ACROSIN.

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.

SEQUENCE FROM N.A.

TISSUE-LIVER;

KEIME S.;
Submitted (Aug-1992) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

RC TISSUE-LIVER;
RX MEDLINE: 96348713.
RA ADHAM I.M., KREMLING H., NIETTER S., ZIMMERMANN S., HUMMEL M.,
RA SCHROETER U., ENGEL W.;
RT "The structures of the bovine and porcine proacrosin genes and their
RT conservation among mammals."

RL Biol. Chem. Hoppe-Seyler 377:261-265(1996).
DR EMBL: X68212; CAA48294.1; -.
DR HSP; P00763; IDPO.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.

DR HSP; P00763; IDPO.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
SQ SEQUENCE 377 AA; 41722 MW; E646321D CRC32;

Query Match 29.58; Score 558; DB 6; Length 377;
Best Local Similarity 37.9%; Pred. No. 2.88e-110;
Matches 96; Conservative 56; Mismatches 82; Indels 19; Gaps 15;

DB 3 IIGGQAGSWMWVMSLOFTYHNNRYVCGLLNAHWLTAHCFRIKKKVDWRL 62

QY 578 IIGKNSLRGMPWVSLKSSHGGRLLCGATLLSSCWLTAAHCFR-RYGNST-R- 633

DB 63 IFGAEKVENGSNKPVPLOERY-VEKIIHEKYSASSEANDIALIKITPPVI-CGPF-- 118

QY 634 SYAVR-VG-DYHTLVPEEFEEELGVQIYIHREYRDRSDYDALVRLGPEEQCARFSS 691

DB 119 -IGPGLPFRAGPRPVYPCVWAGGFLLENARRTSVPLQEAHVDLIDLDLONSTRWTN 177

QY 692 HVLPAQLPLMRERPORTASNCYITGNG--DTGRAYSRILQQAIPLLPRFCEE-R-YK 746

DB 178 GRIKRSINVCAGY-PEKG-IDTCGGSGGGLMKDSVNSVYVVGISWVGCSRRARPGV 235

QY 747 GRTGMLCAGNHEHKKRVDSGSGGGLMC-ERPEGSWVYGVTSWVGCGVDRSPGV 805

DB 236 YTSYSLNMIAS 248

QY 806 YTKVSAFVPMIKS 818

RESULT 6 PRELIMINARY: PRT: 1042 AA.

AC 09Y505;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CORIN.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RC TISSUE=HEART;

SEQUENCE FROM N.A.

SEQUENCE FROM N.A.

RX MEDLINE: 99262646.
RA YAN W., SHENG N., SENO M., MORSE J., WU Q.;
RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
RT from human heart."

RT J. Biol. Chem. 274:14926-14935(1999).
DR EMBL: AF133845; AAD31850.1; -.
DR PROSITE; PS01209; LDLRA_1; 6.
KW Glycoprotein.

SQ SEQUENCE 1042 AA; 116565 MW; FE3227D1 CRC32;

Query Match 29.38; Score 554; DB 4; Length 1042;
Best Local Similarity 40.5%; Pred. No. 3.12e-109;
Matches 98; Conservative 43; Mismatches 87; Indels 14; Gaps 10;

DB 802 IIGGRTSRGPRMPCSLD--SEPS-GHI-CGCVLAKKWTVAHCFGRENA-VMKY 856

QY 578 IIGKNSLRGMPWVSLKSSHGGRLLCGATLLSSCWLTAAHCFRIKYSTSYAV 637

DB 857 VGINNLDHPSVFMQTRFKTIIHPRYSRAVDYDISTVE--SEDISE--TGVRPVC 912

QY 638 RVGDYHTLVPEEFEEELGVQIYIHREYRDRSDYDALVRLGPEEQCARFSSHVLPA 697

DB 913 LPR-PEQMLEPDTCTCTGCHGKMKPFKLOGEVRIITSLHCOSYFPMKITTMTICA 971

QY 698 LPLMRERPORTASNCYITGNGDTRAYSRILQQAIPLLPRFCEEYRGR-FTGRLCA 756

DB 972 G--YESGTYDSCGDSGGLVCEKPGCRWTLFGITSGSVCFSKYVPGVSNVSYFEM 1029

QY 757 GNLHEKRRVDSGSGGLMCEPESWVYGVTSWVGCGVKD-SFGVTKVSAFVPM 815

DB 1030 IK 1031

QY 816 IK 817

RESULT 7 PRELIMINARY: PRT: 421 AA.

AC 060491;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PREPROACROSIN PRECURSOR (FRAGMENT).

OC Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.

RC SEQUENCE FROM N.A.

RA STRAIN=HARTLEY; TISSUE=TESTIS;
RA GERTON G.L., HOFF H.B., BABA T.;
RA Submitted (May-1992) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z12153; CAA78137.1; -.
DR HSP; P00763; IDPO.
DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00089; trypsin; 1.

QY 692 HYLPACLPMPRRPCKATSNCTITGMCDT-GRA-Y-SRTIQQAIPLLPKRCE--ERYK 746
DB 216 GRWMTNVCAGY-PEGK-IDTCGGDSGGPLMCDMANSPVVVYGITSMGVCARAKRPGI 273
QY 747 GFTGMLCAGNLHHRKRVDSGGDSGGLMC-ERPESWVYGVTSWYGCYKDSGPV 805
DB 274 YATATWYLDWIAS 286
QY 806 YTKVSAFVPWIKS 818

RESULT 8
ID 015844 PRELIMINARY: PRT: 411 AA.
AC 015844:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
RT UROKINASE-TYPE PLASMINOGEN ACTIVATOR (FRAGMENT).
OC Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 86056954.
RA NAGAI M., HIRAMATSU R., KANEDA T., HAYASUKE N., ARIMURA H.,
RA NISHIDA M., SOYAMA T.,
RT "Molecular cloning of cDNA coding for human preprourokinase."
RL Gene 36:183-188(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96186279.
RA YOSHIMOTO M., USHIYAMA Y., SAKAI M., TAMAKI S., HARA H., TAKAHASHI K.,
RA SAWASAKI Y., HANADA K.,
RT "Characterization of single chain urokinase-type plasminogen activator
with a novel amino-acid substitution in the kringle structure."
RL Biochim. Biophys. Acta 1293:83-89(1996).
DR HSSP: P00749; 10RK.
DR PFAM: PF00051; kringle; 1.
DR PFAM: PF00089; trypsin; 1.
KW Serine protease.
FT NON_TER 1 1 L -> P (IN REF. 1).
FT CONFLICT 121 121
SQ SEQUENCE 411 AA; 46384 MW; 254CBBDE CRC32;

Query Match 28.7%; Score 543; DB 4; Length 411;
Best Local Similarity 36.9%; Pred. No. 2,17e-106;
Matches 93; Conservative 60; Mismatches 85; Indels 14; Gaps 13;
159 IIGGFTTLENOPWFAIYRRHGGSVTVYCGSLISPCWVISATHCFTDYPKE-DYIV 217
159 IIGGFTTLENOPWFAIYRRHGGSVTVYCGSLISPCWVISATHCFTDYPKE-DYIV 217
578 IIGGKNSLGGPMPWQSLRLKSHGDRLLCGATLLSCVLTAAHCFKRYGSTRSYAV 637
DB 218 YLGRRLNSNTGEMKFEVENLLHKDYADTLAHHNDIALKIRKEGCAQPSRTIOT 277
QY 638 RRGDYHTLVPEEFEEIGVOQIVIHREYRPDR-SDY-DIALVRLQGPEDCCAFSSHVP 695
DB 278 ICLPSMYND-PO-FGSCITGFGKENSVDLYPEOLKMTVYLHRECOQHYGSEV 335
QY 696 ACLP-LMREPRKTSNCTITGMG-DTGRAY--SRTLQQAIPLLKRCFE-RYGR-F 749
DB 336 TTKMLCAD-PQWK-TDSCGDSGGPLVCSLOGRM-TLTGIVSWGRCALDKRPVYTRV 392
QY 750 TGRMLCAGNLHHRKRVDSGGDSGGLMCEPESWVYGVTSWYGCYKDSGPVYTKV 809

RESULT 9
ID 014520 PRELIMINARY: PRT: 560 AA.
AC 014520:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
RT HGF ACTIVATOR LIKE PROTEIN.
OC Homo sapiens (Human)
OC Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RP KITAMURA N.;
RT Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE: 96425001.
RA CHOI-MIRA N.H., TOBE T., SUMIYA J., NAKANO Y., SANO Y., MAZDA T.,
RA TOMITA M.;
RT "Purification and characterization of a novel hyaluronan-binding
protein (PHBP) from human plasma: it has three EGF, a kringle and a
serine protease domain, similar to hepatocyte growth factor
activator."
RL J. Biochem. 119:1157-1165(1996).
DR EMBL: D49742; BAA08576.1; -.
DR EMBL: S83182; AAB46909.1; -.
DR HSSP: P00763; IDPO.
DR PFAM: PF00008; EGF; 3.
DR PFAM: PF00051; kringle; 1.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PR00018; KRINGLE.
DR PRINTS: PR00722; CHYMOTRYPSIN.
SQ SEQUENCE 560 AA; 62671 MW; 4AC81907 CRC32;

Query Match 28.6%; Score 541; DB 4; Length 560;
Best Local Similarity 41.8%; Pred. No. 7,12e-106;
Matches 99; Conservative 45; Mismatches 73; Indels 20; Gaps 14;
332 QSLPLTISMPQGH-CGALLHPCWVLTAAHC-T-DIKTHLKVGLD-ODLKEEFH 386
QY 592 QVSLRLKSHGDRLLCGATLLSCVLTAAHCFKRYGSTRSYAVRGDYHTLVPEEF 651
DB 387 EOSFREKIFKYSHYNEREIHNDIALILKLPVGHCLAESKYKYICLP---DGSFPS 443
QY 652 EE-IGVOQIVIHREY--RPDRSDYDIALVRLQGPEDCCAFSSHVPCLPMRRPPOKT 708
DB 444 GSECHISGCVETGKG-SROLDAKVKLIANTLNSRQYDHMDIDSMICAGNLQKQ 502
QY 709 ASNCYITGMG-DTGRAYRSLQQAIPLLPKRCEER--YGRGTGMLCAGNLHHRK 764
DB 503 -DTCCGDSGGPLTCEKDQ-TTYVYGVSWGLECG-K-RPGVYTVYTKFLNMIKATIK 555
QY 765 VPSCCGDSGGPLMCEPESWVYGVTSWYGCYKDSGPVYTKVSAFVPWIKSVTK 821

RESULT 10
ID 092319 PRELIMINARY: PRT: 1113 AA.
AC 092319:
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
RT LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN 4.
OC Mus musculus (Mouse).
OC Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 96429596.
RA TOMITA Y., KIM D.-H., MAGOORI K., FUJINO T., YAMAMOTO T.T.;
RT "A novel low-density lipoprotein receptor-related protein with type II
membrane protein-like structure is abundant in heart."
RL J. Biochem. 124:784-789(1998).
DR EMBL: AB013874; BAA34371.1; -.
DR HSSP: P00763; IDPO.
DR PROSITE: PS01209; LDLRA_1; 6.

KW Glycoprotein.
SQ SEQUENCE 1113 AA; 122984 MW; ABADC31E CRC32;

Query Match 28.6%; Score 542; DB 11; Length 1113;
Best Local Similarity 40.7%; Pred. No. 3,936-106;
Matches 98; Conservative 39; Mismatches 90; Indels 14; Gaps 11;

Db 869 ILGRTSRGRMPWCSLQ--SEPS-GHT-CGCVLAKKWLTVANCFEGREDAD-VWKY 933

578 IIGKNSLRGMPWQVSLRLKSSHGDRLLCGATLLSSCVLTAACFKRYGSTRSYAV 637

524 VFGINLIDHPSGMQRFVYVITLLHPYRSRAVVDYDSVLEL--SDI-NE-TSYRPPV 979

638 RVGDYHTLVPDEFEEELGVQVQIVIHREYRDRSDYDIALVRLQGEQCARFSSHVLPAC 697

980 LP-SPEEYLEDPTCYITGWHGNNKMFELQGEVRIIPLEOCQSYFDKRTITNMICA 1038

698 LPLMRERPKTASNCYITGDDGGRAYSRLQQAIPILPKRCFEYKGR-FTGRMLCA 756

1039 G--YESGTVDSGMDSGPLVCERPGGQWTLFGLTWSGVSFSGKVLGPGVYSVSVFVGW 1096

757 GNLEHHRVDSGCGDSGGLMCEPGEVSVVYGVTSWGYGCVKD-SPGYITKVSFAFVP 815

Db 1097 I 1097

QY 816 I 816

RESULT 11 PRELIMINARY; PRT; 806 AA.

AC 018783; 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)

DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)

DE PLASMINOGEN.

OS Macropus eugenii (Tamar wallaby).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Metatheria; Diprotodontia; Macropodidae; Macropus.

RC TISSUE=LIVER;

RA LAMN R.M., SCHWARTZ K., PATTHY L.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF012297; AAB65760.1; -

DR HSSP; P00747; 5HPG.

PFAM; PF00051; kringie; 5.

PFAM; PF00089; trypsin; 1.

PRINTS; PR00018; KRINGLE.

PRINTS; PR00722; CHYMOTRYPSIN.

SEQUENCE 806 AA; 90980 MW; 4371C1C6 CRC32;

Query Match 28.3%; Score 535; DB 6; Length 806;

Best Local Similarity 40.9%; Pred. No. 2,516-104;

Matches 101; Conservative 53; Mismatches 69; Indels 22; Gaps 17;

Db 577 IYGGCYAOPHSWQISLRT--FGE-HF-CGGTLAPQWVLAACLER-SOMPGAYKY 631

578 IIGKNSLRGMPWQVSLRLKSSHGDRLLCGATLLSSCVLTAACFKRYGSTRSYAV 637

632 ILGLHREVNESQELGVSR-L-F-KG--P-LAA-DIALIKLRPA--A-INDKYIPAC 661

638 RVGDYHTLVPDEFEEELGVQVQIVIHREYRDRSDYDIALVRLQGEQCARFSSHVLPAC 697

682 LP-SQDFWMPDRFLCHVTGMDRGSTSPRGLKQASLPVDNVCNRHEHLNRVSTEL 740

698 LPLMRERPKTASNCYITGDDGGRAYSRL-QQAIPILPKRCFEYKGR-FTGRMLCA 754

741 CAGHLVG--RGDSGCGSGGPLIC-FEDDKYVLAQVTSWGLGAPRNKPGVYRVSRYIS 797

755 CAGNLEHHRVDSGCGDSGGLMCEPGEVSVVYGVTSWGYGCVKDSPPGYITKVSFAFVP 814

798 WIEDVAK 804

QY 815 WIKSVTK 821

RESULT 12 PRELIMINARY; PRT; 431 AA.

AC 016618; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)

DE UROKINASE PRECURSOR (EC 3.4.99.26).

GN PLAU.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RP SEQUENCE FROM N.A.

RA MEDLINE; 85203359.

RA JACOBS P., CRAVADOR A., LORIAN R., BROCKLEY F., COLAU B., CHUCHANA P.,

RA VAN ELSEN A., HERZOG A., BOLLER A.;

RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of

RT human preprothrombin cDNA."

RL DNA 4:139-146(1985).

DR EMBL; X02760; CAA26535.1; -

DR HSSP; P00749; IURK.

DR PFAM; PF00051; kringie; 1.

DR PFAM; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

KW Signal; Serine protease.

FT SIGNAL 1 20

FT CHAIN 21 431

SQ SEQUENCE 431 AA; 48664 MW; 988C805F CRC32;

Query Match 28.0%; Score 530; DB 4; Length 431;

Best Local Similarity 36.5%; Pred. No. 4,886-103;

Matches 92; Conservative 60; Mismatches 86; Indels 14; Gaps 13;

Db 179 IIGGEFTTENOPWFAIYRRHRGSGVYVCGSLISPCWVISAHCFIDYPRKE-DIYV 237

578 IIGKNSLRGMPWQVSLRLKSSHGDRLLCGATLLSSCVLTAACFKRYGSTRSYAV 637

238 YLGRSLNNTGEMKEFEVNLILHKDYADILAHNDIALKIRKEGRCAOPSTIQT 297

638 RVGDYHTLVPDEFEEELGVQVQIVIHREYRDR-SDY-DIALVRLQGEQCARFSSHVLP 695

298 ICDPSMYND-PQ-FGTSCITGFGKENSVDYLPEDLKYTVYLISHRCOOPHYGSEV 355

696 ACLP-LMRERPKTASNCYITGDDGGRAYSRL-QQAIPILPKRCFEYKGR-FTGRMLCA 754

413 SHLEPWIRSHTK 424

810 SAFVPMWIKSVTK 821

RESULT 13 PRELIMINARY; PRT; 643 AA.

AC 097506; 01-MAY-1999 (TREMblrel. 10, Created)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)

DE KALLIKREIN

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

RP SEQUENCE FROM N.A.

RA TAKAHASHI T., KIMURA A., OKIMURA H., HAMABATA T.;

RT "Porcine liver plasma kallikrein."

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

 W O R L D
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:19:00 2000; MasPar time 12.76 Seconds
 454.820 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-147-947-6
 Description: (578-822) from US09147947A.ppt (1 of 6)
 Perfect Score: 1892
 Sequence: 1 IIGKNSLRGMPQVSLRL.....PGVYTKVSAPVPMIKSVTKL 245

Scoring table: PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 32.789; Variance 126.286; scale 0.260

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	DB ID	Description	Pred. No.
1892	100.0	822	1 W99087	Human serine protease	9.99e-193
1892	100.0	875	1 W83361	Human neurotrophin	9.99e-193
1892	100.0	92.6	1 W83362	Mouse neurotrophin	3.43e-177
1752	92.6	761	1 W99088	Mouse serine protease	3.43e-177
561	29.1	522	1 R13921	Delta (466-470) tpa va	1.35e-45
559	29.5	522	1 R13920	Delta (466-470) tpa va	2.23e-46
554	29.3	356	1 W46917	Amino acid sequence of	7.65e-46
552	29.2	395	1 R47902	Pro-urokinase derivative	9.78e-46
552	29.2	434	1 R13922	Delta(92-179) and delt	1.25e-45
553	29.2	522	1 R13918	Delta (466-470) tpa va	9.78e-46
552	29.2	522	1 R14486	Delta (466-470) tpa va	1.25e-45
551	29.1	412	1 R05116	UK-SI as encoded by ps	1.60e-45
551	29.1	522	1 R13919	Delta (466-470) tpa va	1.60e-45
548	29.0	378	1 W13635	Human prourokinase var	3.35e-45
548	29.0	389	1 W13636	Human prourokinase var	3.35e-45
548	29.0	401	1 W13637	Human prourokinase var	3.35e-45
548	29.0	411	1 W13634	Human native prourokin	3.35e-45
548	29.0	411	1 R47958	PUR S26T.	3.35e-45
548	29.0	411	1 R47959	PUR G32P.	3.35e-45
548	29.0	411	1 R47963	PUR G53A.	3.35e-45
548	29.0	411	1 R47964	PUR N32P G38K.	3.35e-45
548	29.0	411	1 R47957	PUR Y24A.	3.35e-45
548	29.0	411	1 R40225	PUR.	3.35e-45

24	548	29.0	411	1 R47966	PUR G16A.	3.35e-45
25	548	29.0	411	1 R47960	PUR P34A.	3.35e-45
26	548	29.0	411	1 R47965	PUR N32P G38K G39K.	3.35e-45
27	548	29.0	411	1 R47956	PUR G16A G17A.	3.35e-45
28	548	29.0	411	1 R63003	Pro-urokinase mutant S	3.35e-45
29	548	29.0	411	1 R47961	PUR G38A.	3.35e-45
30	548	29.0	411	1 R47962	PUR G38A G39A.	3.35e-45
31	547	28.9	411	1 R63008	Pro-urokinase mutant S	4.28e-45
32	547	28.9	411	1 R63009	Pro-urokinase mutant H	4.28e-45
33	547	28.9	411	1 R63007	Pro-urokinase mutant S	4.28e-45
34	546	28.9	559	1 R13263	JM1-229 cell line t-pa	5.47e-45
35	545	28.8	390	1 R22504	[GARSQ]-[Plasminogen	7.00e-45
36	545	28.8	411	1 R80428	Sequence of mutants of	7.00e-45
37	545	28.8	411	1 R04185	Plasminogen activator	7.00e-45
38	545	28.8	411	1 R63002	Pro-urokinase mutant S	7.00e-45
39	544	28.8	411	1 R63000	Pro-urokinase mutant S	8.95e-45
40	544	28.8	411	1 R62997	Pro-urokinase mutant G	8.95e-45
41	544	28.8	411	1 P93504	Sequence of human pro-	8.95e-45
42	544	28.8	412	1 R10334	Recombinant single-cha	8.95e-45
43	545	28.8	430	1 W24578	Inhibitor resistant ur	7.00e-45
44	545	28.8	431	1 P94764	Non-glycosylated prout	7.00e-45
45	545	28.8	476	1 R22503	[GARSQ]-[Plasminogen	7.00e-45

ALIGNMENTS

RESULT 1
 ID W99087 standard; Protein: 822 AA.
 AC W99087:
 DT 13-MAY-1999 (first entry)
 DE Human serine protease BSSP-3.
 KW Serine protease; BSSP-3; brain tissue.
 OS Homo sapiens.
 PN W09905290-A1.
 PD 04-FEB-1999.
 PE 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SUNR) SUNTORY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR N-PSDB: X19024.
 PT New serine protease expressed in brain tissue - used in screening
 for potential serine protease inhibitors for drug use
 PS Claim 1; Page 61-65; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 is isolated from human brain tissue. Transforms may be used to
 produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 822 AA:

Query Match	100.0%	Score 1892	DB 1: Length 822
Best Local Similarity 100.0%;	Pred. No. 9.99e-193;		
Matches 245; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Db	578 IIGKNSLRGMPQVSLRLKSSHDGRLLCGATLLSCGWTAAHCFKYNSTRSYAV	637	
QY	578 IIGKNSLRGMPQVSLRLKSSHDGRLLCGATLLSCGWTAAHCFKYNSTRSYAV	637	
Db	638 RVGDYHTLVPEEFEEIGVQIVIHREYRPDRSDYDIALVRLQGEEOCARFSSHYLPAC	697	
QY	638 RVGDYHTLVPEEFEEIGVQIVIHREYRPDRSDYDIALVRLQGEEOCARFSSHYLPAC	697	
Db	698 LPLMERPOKTSNCTYITGMDTGRAYSRTLOQAALPLPRFCERKRGRTGMLAG	757	
QY	698 LPLMERPOKTSNCTYITGMDTGRAYSRTLOQAALPLPRFCERKRGRTGMLAG	757	
Db	758 NLHEHRRVDSGCGDGLMCEPESWVYGVTSNGYCGCYKDGSPGYTKVSAPVPMIK	817	
QY	758 NLHEHRRVDSGCGDGLMCEPESWVYGVTSNGYCGCYKDGSPGYTKVSAPVPMIK	817	
Db	818 SVTKL 822		
QY	818 SVTKL 822		

QY 818 SVTKL 822

RESULT 2

ID W83361 standard; Protein: 875 AA.

AC W83361:

DT 17-FEB-1999 (first entry)

DE Human neurotrophin.

KW Human; neurotrophin; tumour inhibition; neurological disease;

KW lung disease; gene therapy; drug development; stroke; brain injury;

KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;

KW epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;

KW emphysema; bronchitis.

OS Homo sapiens.

PN W09849322-A1.

PD 05-NOV-1998.

PF 24-APR-1998; IB0625.

PT 26-APR-1997; CH-000966.

PS (SOND/) SONDEREGGER P.

PT Sanderegger P.

PI: 99-009438/01.

N-PSDB: V72589.

PT New human and murine neurotrophin - used, e.g. for inhibiting

tumours, treatment of neurological or lung disease, including by

gene therapy and in drug development

PS Claim 1; Page 20-24; 50pp; English.

CC The present sequence represents human neurotrophin. Neurotrophin proteins

and polynucleotides can be used: (i) to inhibit tumours, including

metastases, e.g. of brain or retina; (ii) to minimise tissue damage

caused by stroke or brain injury (having a protective effect on the

penumbra zone); (iii) to treat or prevent neurodegeneration,

neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

increase survival of damaged neurons (e.g. in cases of hypoxia,

ischaemia, nerve transection) and to stimulate regeneration and/or

restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

degeneration or neurodegeneration); (vi) to prevent apoptosis (or other

causes of cell death) in the nervous system; (vii) to regenerate brain

and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

performance, including learning and memory; (x) to treat or prevent a

wide range of psychiatric disorders; and (xi) to treat brain or lung

injury associated with protease expression (specifically emphysema or

bronchitis).

CC Sequence 875 AA:

Query Match 100.0%; Score 1892; DB 1; Length 875;

Best Local Similarity 100.0%; Pred. No. 9,99e-193;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 631 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSSTSYAV 630

QY 578 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSSTSYAV 637

DB 691 RRGVDHTLVPEEFEEIEIGVOQVIHREYRPDRSDYDIALVRLGPEOCRRFSSHVLPAC 750

QY 638 RRGVDHTLVPEEFEEIEIGVOQVIHREYRPDRSDYDIALVRLGPEOCRRFSSHVLPAC 697

DB 751 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAIPLPKRFCEERYKRGFTGRMLCAG 810

QY 698 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAIPLPKRFCEERYKRGFTGRMLCAG 757

DB 811 NLHEKRVDSGCGDGGPLMCEPGESEWVYGVTSWYGGCGVADPGVYTKVSAFVPMIK 870

QY 758 NLHEKRVDSGCGDGGPLMCEPGESEWVYGVTSWYGGCGVADPGVYTKVSAFVPMIK 817

DB 871 SVTKL 875

QY 818 SVTKL 822

RESULT 3

ID W83362 standard; Protein: 761 AA.

AC W83362:

DT 17-FEB-1999 (first entry)

DE Human neurotrophin.

KW Human; neurotrophin; tumour inhibition; neurological disease;

KW lung disease; gene therapy; drug development; stroke; brain injury;

KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;

KW epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;

KW emphysema; bronchitis.

OS Homo sapiens.

PN W09849322-A1.

PD 05-NOV-1998.

PF 24-APR-1998; IB0625.

PT 26-APR-1997; CH-000966.

PS (SOND/) SONDEREGGER P.

PT Sanderegger P.

PI: 99-009438/01.

N-PSDB: V72589.

PT New human and murine neurotrophin - used, e.g. for inhibiting

tumours, treatment of neurological or lung disease, including by

gene therapy and in drug development

PS Claim 1; Page 20-24; 50pp; English.

CC The present sequence represents human neurotrophin. Neurotrophin proteins

and polynucleotides can be used: (i) to inhibit tumours, including

metastases, e.g. of brain or retina; (ii) to minimise tissue damage

caused by stroke or brain injury (having a protective effect on the

penumbra zone); (iii) to treat or prevent neurodegeneration,

neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

increase survival of damaged neurons (e.g. in cases of hypoxia,

ischaemia, nerve transection) and to stimulate regeneration and/or

restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

degeneration or neurodegeneration); (vi) to prevent apoptosis (or other

causes of cell death) in the nervous system; (vii) to regenerate brain

and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

performance, including learning and memory; (x) to treat or prevent a

wide range of psychiatric disorders; and (xi) to treat brain or lung

injury associated with protease expression (specifically emphysema or

bronchitis).

CC Sequence 875 AA:

Query Match 100.0%; Score 1892; DB 1; Length 875;

Best Local Similarity 100.0%; Pred. No. 9,99e-193;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 631 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSSTSYAV 630

QY 578 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSSTSYAV 637

DB 691 RRGVDHTLVPEEFEEIEIGVOQVIHREYRPDRSDYDIALVRLGPEOCRRFSSHVLPAC 750

QY 638 RRGVDHTLVPEEFEEIEIGVOQVIHREYRPDRSDYDIALVRLGPEOCRRFSSHVLPAC 697

DB 751 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAIPLPKRFCEERYKRGFTGRMLCAG 810

QY 698 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAIPLPKRFCEERYKRGFTGRMLCAG 757

DB 811 NLHEKRVDSGCGDGGPLMCEPGESEWVYGVTSWYGGCGVADPGVYTKVSAFVPMIK 870

QY 758 NLHEKRVDSGCGDGGPLMCEPGESEWVYGVTSWYGGCGVADPGVYTKVSAFVPMIK 817

DB 871 SVTKL 875

QY 818 SVTKL 822

RESULT 4

ID W93088 standard; Protein: 761 AA.

AC W93088:

DT 13-MAY-1999 (first entry)

DE Mouse serine protease BSSP-3.

KW Mouse; serine protease; BSSP-3; brain tissue.

KW lung disease; gene therapy; drug development; stroke; brain injury;

KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;

KW epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;

KW emphysema; bronchitis.

OS Mus musculus.

PN W09905290-A1.

PD 04-FEB-1999.

PF 24-JUL-1997; JP-213969.

PT 24-JUL-1997; JP-213969.

PS (SOND/) SONDEREGGER P.

PT Sanderegger P.

PI: 99-009438/01.

N-PSDB: V72590.

PT New human and murine neurotrophin - used, e.g. for inhibiting

tumours, treatment of neurological or lung disease, including by

gene therapy and in drug development

PS Claim 1; Page 29-32; 50pp; English.

CC The present sequence represents mouse neurotrophin. Neurotrophin proteins

and polynucleotides can be used: (i) to inhibit tumours, including

metastases, e.g. of brain or retina; (ii) to minimise tissue damage

caused by stroke or brain injury (having a protective effect on the

penumbra zone); (iii) to treat or prevent neurodegeneration,

neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

increase survival of damaged neurons (e.g. in cases of hypoxia,

ischaemia, nerve transection) and to stimulate regeneration and/or

restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

degeneration or neurodegeneration); (vi) to prevent apoptosis (or other

causes of cell death) in the nervous system; (vii) to regenerate brain

and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

performance, including learning and memory; (x) to treat or prevent a

wide range of psychiatric disorders; and (xi) to treat brain or lung

injury associated with protease expression (specifically emphysema or

bronchitis).

CC Sequence 761 AA:

Query Match 92.6%; Score 1752; DB 1; Length 761;

Best Local Similarity 89.8%; Pred. No. 3,43e-177;

Matches 220; Conservative 17; Mismatches 8; Indels 0; Gaps 0;

DB 517 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSSTSYAV 576

QY 578 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSSTSYAV 637

DB 691 RRGVDHTLVPEEFEEIEIGVOQVIHREYRPDRSDYDIALVRLGPEOCRRFSSHVLPAC 750

QY 638 RRGVDHTLVPEEFEEIEIGVOQVIHREYRPDRSDYDIALVRLGPEOCRRFSSHVLPAC 697

DB 751 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAIPLPKRFCEERYKRGFTGRMLCAG 810

QY 698 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAIPLPKRFCEERYKRGFTGRMLCAG 757

DB 811 NLHEKRVDSGCGDGGPLMCEPGESEWVYGVTSWYGGCGVADPGVYTKVSAFVPMIK 870

QY 758 NLHEKRVDSGCGDGGPLMCEPGESEWVYGVTSWYGGCGVADPGVYTKVSAFVPMIK 817

DB 871 SVTKL 875

QY 818 SVTKL 822

OS Synthetic.
 FH Key Location/Qualifiers
 FT region 1..10 "residues 1-10 of native hPUK"
 FT region /note= "residues 1-10 of native hPUK"
 FT region 11..389
 FT region /note= "residues 33-411 of native hPUK"
 FT misc_difference 155
 FT /note= "corresponds to TAC codon"
 PN EP-398361-A.
 PD 22-NOV-1990.
 PE 18-MAY-1990: 109472.
 PR 03-JUN-1986: JP-156936.
 PR 18-FEB-1987: JP-036495.
 PR 18-MAY-1989: JP-126434.
 PR 18-MAY-1989: JP-126433.
 PR 22-FEB-1990: JP-042020.
 (GREG, GREEN CROSS CORP.
 Airmura H, Amatsuji Y, Hirose M, Kasai S, Kawabe H;
 Morita M, Tanabe T;
 DR MPI; 90-350146/47.
 DR N-PSDB; T61673.
 PT Human pro-urokinase variants - deficient in loop regions of
 PT epidermal growth factor, showing long blood half-life, as
 PT fibrinolytic agent
 PS Claim 6; Page -; 22pp; English.
 CC New variants of human prourokinase (hPUK) comprise a hPUK deficient
 CC in (i) at least part of the first loop region of the epidermal growth
 CC factor (EGF) domain; (ii) at least part of the first loop and at
 CC least part of the second loop; or (iii) at least part of the third
 CC loop. The hPUK variants show an increased blood half-life comparable
 CC to that of the whole EGF domain-deficient hPUK variant and urokinase
 CC while retaining the same properties as those of hPUK. They have
 CC potent thrombolytic activity and very little tendency to cause
 CC spontaneous bleeding. The present sequence represents a specific
 CC variant of hPUK which lacks loops 1 and 2 of the EGF domain; the
 CC sequence does not appear in the specification and has been created
 CC using the wild-type hPUK sequence and the junction sequence after
 CC deletion, both of which are given (in Fig 1 and on page 8,
 CC respectively).
 SQ Sequence 389 AA;

Query Match 29.0%; Score 548; DB 1; Length 389;
 Best Local Similarity 36.9%; Pred. No. 3.35e-45;
 Matches 93; Conservative 60; Mismatches 85; Indels 14; Gaps 13;

Db 137 IIGGEFTIEQPPFAITRRHRGGSVTVYVCGSLISPCWISATHCIDPKKE-DYIV 195
 578 IIGKNSLRGCMQVQVSLRKSXSHGDRLLCGATLLSSCWLTAAHCFKRYGNSTRSYAV 637
 196 YLGRSRINSNTOGEMKEVENTLHKDYADTLAHNDIALKIRKRGCAQPSRTIQI 255
 QY 638 RVGYHTLVPEFEFEICVQIVIHREYRPR-SDY-DIALVRLQGPEDQCARFSSHYLP 695
 Db 256 ICLPSMTND-PQ-FGTSGEITFGKENSTDYLYPDLKMTVVKLISHRECQPHYGSEV 313
 QY 696 ACLP-LWREPOKFTASNCYITGMG-DTGRAY--SRTIQOAAIPLPKFCEE-RYKGR-F 749
 Db 314 TTKMLCAD-PQWK-TDSCGDSGGLVCSLQGM-TLTGIVSGRCALCKDKPGYTRV 370
 QY 750 TGRMLCAGNLHHRKRVSCQDSGGLMCEPREGSWVYGVTSWYGCGVNDSPGYTKV 809
 Db 371 SHPLPWIRSHTK 382
 QY 810 SAFVPIKSVTK 821

Search completed: Mon Mar 13 10:19:14 2000
 Job time : 14 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

In_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:18:23 2000; Maspar time 18.79 Seconds
Tabular output not generated. 615.076 Million cell updates/sec

```

```

Title: >US-09-147-947-6
Description: (578-822) from US09147947A.pep (1 of 6)
Perfect Score: 1892
Sequence: 1 IIGKNSLRGGMQVSLR.....PCVYIKVSAFPWIKSTYTL 245

```

Scoring table: PAM 150

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:      pir62
               1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 44.068; Variance 74.731; scale 0.590

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

T	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	1752	92.6	761	2	JC5759		brain-specific serine	0.00e+00
2	625	33.0	625	1	KFNU1		coagulation factor XI	1.09e-11
3	600	31.7	638	1	KQMSPL		plasma kallikrein (EC	1.09e-11
4	594	31.4	638	1	KORP1		plasma kallikrein (EC	2.92e-10
5	594	31.4	638	1	KOHUP		plasma kallikrein (EC	2.92e-10
6	574	30.3	415	1	A34170		acrosin (EC 3.4.21.10)	1.65e-10
7	560	29.6	418	2	A33344		acrosin (EC 3.4.21.10)	3.43e-10
8	557	29.4	437	2	S14047		acrosin (EC 3.4.21.10)	1.76e-10
9	555	29.3	431	2	S47538		acrosin (EC 3.4.21.10)	5.24e-10
10	547	28.9	436	2	JX0172		acrosin (EC 3.4.21.10)	4.08e-08
11	546	28.9	459	1	A35029		t-plasminogen activat	7.03e-07
12	545	28.8	421	2	S29599		acrosin (EC 3.4.21.10)	1.21e-97
13	545	28.8	431	1	UKRUU		u-plasminogen activat	1.21e-97
14	541	28.6	560	1	UC4795		plasma hyaluronan-bin	1.07e-86
15	542	28.6	1113	2	JF0315		low-density lipoprote	6.19e-97
16	536	28.3	790	1	PLPG		plasmin (EC 3.4.21.7)	1.61e-95
17	527	27.9	559	1	AK2941		t-plasminogen activat	2.14e-93
18	527	27.9	562	1	UKHUT		t-plasminogen activat	8.30e-91
19	516	27.3	558	2	JC5878		plasma hyaluronan-bin	2.45e-90
20	514	27.2	421	1	S11674		acrosin (EC 3.4.21.10)	7.32e-90
21	512	27.1	431	1	UKRAY		u-plasminogen activat	4.21e-90
22	513	27.1	477	1	A34369		t-plasminogen activat	4.21e-90
23	510	27.0	455	2	AB1545		plasmin (EC 3.4.21.7)	2.13e-89

Postage
-947-6-01. rpr
lots 2 are acceptable

24	505	26.7	655	1	A46688	hepatocyte growth fac	3.18e-88
25	504	26.6	442	1	UKP6	u-plasminogen activat	5.47e-88
26	504	26.6	460	2	B61545	plasmin (EC 3.4.21.7)	5.47e-88
27	503	26.6	812	1	PLB0	plasmin (EC 3.4.21.7)	9.38e-88
28	503	26.6	1019	1	A56318	urokinase (EC 3.4.21.7)	9.38e-88
29	501	26.5	431	2	JS0599	t-plasminogen activat	2.76e-87
30	499	26.5	394	2	JS0600	t-plasminogen activat	8.13e-87
31	499	26.4	477	2	JS0598	t-plasminogen activat	8.13e-87
32	499	26.4	477	2	JS0597	t-plasminogen activat	8.13e-87
33	499	26.4	553	2	S45281	coagulation factor XI	8.13e-87
34	489	25.8	433	1	JN0560	u-plasminogen activat	1.79e-84
35	487	25.7	810	1	PLH0	plasmin (EC 3.4.21.7)	5.25e-84
36	484	25.6	812	1	P1M5	plasmin (EC 3.4.21.7)	2.64e-83
37	481	25.4	444	1	A35005	u-plasminogen activat	1.33e-82
38	476	25.2	810	2	B30648	plasmin (EC 3.4.21.7)	1.95e-81
39	474	25.1	264	2	I18136	chymotrypsin-like pro	5.72e-81
40	474	25.1	613	2	I51658	complement C3b/C4b in	5.72e-81
41	472	24.9	433	1	UKM5	u-plasminogen activat	1.67e-80
42	472	24.9	603	2	S28941	coagulation factor XI	1.67e-80
43	472	24.8	285	2	T15451	hypothetical protein	4.90e-80
44	467	24.7	1034	2	A53663	enteropeptidase (EC 3	2.45e-79
45	463	24.5	810	2	I46260	plasmin (EC 3.4.21.7)	2.09e-78

ALIGNMENTS

```

RESULT      1
ENTRY       JC5759      #type complete
TITLE       brain-specific serine proteinase (EC 3.4.21.-) - mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change
            17-Mar-1999
ACCESSIONS  JC5759
REFERENCE   JC5759
AUTHORS     Yamamura, Y.; Yamashiro, K.; Tsunoka, N.; Nakazato, H.;
            Tsujimura, A.; Yamaguchi, N.
            Biochem. Biophys. Res. Commun. (1997) 233:386-392
            Molecular cloning of a novel brain-specific serine protease
            with a kingle-like structure and three scavenger receptor
            cysteine-rich motifs.
            Cysteine-rich motifs.
#cross-references M01D:98008848
#accession      JC5759
#molecule_type mRNA
#residues       1-761 #label YAM
#cross-references DDBJ:D89871
#experimental_source brain
CLASSIFICATION #superfamily trypsin homology; scavenger receptor
               cysteine-rich domain homology
KEYWORDS       glycoprotein; hydrolase; serine proteinase
FEATURE
85-157         #domain kingle-like #status predicted #label KRI\
163-266        #domain scavenger receptor cysteine-rich domain homology
               #label SRC7\
166-266,273-372,
386-486        #domain scavenger receptor cysteine-rich #status
               predicted #label SRC\
513-516        #domain furin binding #status predicted #label FRB\
517-755        #domain trypsin homology #label TRY\
93,521,569     #binding_site carbohydrate (Asn) (covalent) #status
               predicted\
562,612,711    #active_site His, Asp, Ser #status predicted
SUMMARY      #length 761 #molecular_weight 84136 #checksum 5449

Query Match      92.6%; Score 1752; DB 2; Length 761;
Best Local Similarity 89.8%; Pred. No. 0.00e+00;
Matches 220; Conservative 17; Mismatches 8; Indels 0; Gaps 0;

Db 517 IIGGNSLRGAWQMSILRSAGHGRLGATLLSSCWYLAHCKRRKGNRSRYAV 576
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
OY 578 IIGGKNSLRGAWPMOYSLTKSSHGDRLLCATLLSSCWYLAHCKRRKGNSTRYAV 637
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 577 RVGDYHTLVPPEEFDEIGVQAVIHRNRYPRDSVDIALVRLQGGECARLSTHYVPAC 636
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

QY 638 RYGDYHTLVEPEEEFEGVQOIVIHREYRDRSDYDALVRLQGEPEOCARFSSHVLPAC 697

Db 637 LPLMRERPKTASNCNHTGTGTRAYSRLQQAAPLPRKFKERYKGLFTGRMLCAG 696

QY 698 LPLMRERPKTASNCNHTGTGTRAYSRLQQAAPLPRKFKERYKGLFTGRMLCAG 757

Db 697 NQEDNRVSCODSGPLMCEKPDSEWVYVGYTSMGCGVKTPTGVYTVPAFVPMIK 756

QY 758 NHEHRRVSCODSGPLMCEKPDSEWVYVGYTSMGCGVKTPTGVYTVPAFVPMIK 817

Db 757 SVTSL 761

QY 818 SVTKL 822

RESULT 2

ENTRY #type complete

TITLE coagulation factor Xla (EC 3.4.21.27) precursor - human

SM #formal_name Homo sapiens #common_name man

13-Aug-1986 #sequence_revision 26-May-1994 #text_change

18-Jun-1999

ACCESSIONS A27431: A00920: A37940

REFERENCE A27431

#authors Asakai, R.; Davie, E.W.; Chung, D.W.

#journal Biochemistry (1987) 26:7221-7228

#title Organization of the gene for human factor XI.

#cross-references M0ID:88107663

#accession A27431

#molecule_type DNA

##residues 1-625 ##label ASA

##cross-references GB:M18295

REFERENCE A00920

#authors Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.

#journal Biochemistry (1986) 25:2417-2424

#title Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein.

#cross-references M0ID:86243360

#accession A00920

#molecule_type mRNA

##residues 1-625 ##label FUJ

##cross-references GB:M13142; NID:9182832; PIDN:AA52487.1; PID:9182833

REFERENCE A37940

#authors McMullen, B.A.; Fujikawa, K.; Davie, E.W.

#journal Biochemistry (1991) 30:2056-2060

#title Location of the disulfide bonds in human coagulation factor XI: the presence of tandem apple domains.

#cross-references M0ID:91152017

#accession A37940

#molecule_type protein

##residues 28-33:35-49, 'X', 51-55, 'X', 57-63:70-75, 'X', 77-79:107-109, 'X', 111-112:132-139, 'X', 141-154:163-164, 'X', 166-168:192, 'X', 194-198-199, 'X', 223-228:229, 'X', 231-235, 'X', 237, 248:253-254, 'X', 256-258:280-282, 'X', 284:285-287, 313-316, 'X', 318-319:320-326, 'X', 328-330, 'X', 347-349, 373, 'X', 375:377-379, 'X', 381-383:414-415, 'X', 417-431, 'X', 433-437:486-499, 'X', 501-507:535-548:559, 'X', 561-564 ##label MCM

COMMENT The proenzyme consists of two identical chains linked by one or more disulfide bonds. It is activated by factor Xla (or Xll), which cleaves each chain into a light chain, which contains the active site, and a heavy chain, which associates with high molecular weight (HMW) kininogen.

GENETICS

#gene GDB:F11

##cross-references GDB:119891; OMIM:264900

#map_position 4q35-4q35

#introns 19/1: 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 526/1; 572/3

FUNCTION

#description catalyzes the proteolytic activation of coagulation factor IX

#pathway blood coagulation intrinsic pathway

CLASSIFICATION #superfamily coagulation factor XI; trypsin homology

KEYWORDS blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydrolase; plasma; serine proteinase

FEATURE

1-18

19-387

19-108

109-198

199-288

290-379

388-625

388-618

20-103, 514-581,

571-599

29

46-76, 50-56,

110-193, 136-165,

140-146, 200-283,

226-255, 230-236,

291-374, 317-346,

321-327, 380-500,

416-432, 545-560

90, 126, 353, 450

339

387-388

431, 480, 575

491

SUMMARY

length 625 #molecular-weight 70109 #checksum 9314

Query Match 33.0%; Score 625; DB 1; Length 625;

Best Local Similarity 41.88; Pred. No. 1.19e-116;

Matches 102; Conservative 53; Mismatches 71; Indels 18; Gaps 16;

Db 388 IVGGLASVREGEPMQVYLH-TSPPTQ-RHLCGSLIGNOMILTAHCF--YGVESPRIIR 443

QY 578 IIGGKNLRCGMWQVSLRKSSHGRLCGATLLSSCVLTAACFRYRG-NSIRSA 636

Db 444 VYSGILNQSEIKED-TSFGVDEIITHDQYKMAESGYIALKLETVN-YT-DSQR-P 498

QY 637 VRVGD-YHTLVPEPEEEFEGVQOIVIHREYRDRSDYDALVRLQGEPEOCARFSSHVLP 695

Db 499 ICPLSKGDR-NVYITDCWYIGWYRKLRDKIQNTLOKAKIPLVNTNECCRRYRGKITHK 557

QY 696 ACPLMRERPKTASNCNHTGTG-DRGRA-YSRFLQQAAPLPRKFKERYK-RTGR 752

Db 558 MICAGY-REGCK-DACKGDSGGLPSC-KHNEVWHLGIRSMGCAQORPQYTVNVEY 614

QY 753 MLAGNHEHRRVSCODSGPLMCEKPDSEWVYVGYTSMGCGVKTPTGVYTVPAFVPMIK 812

Db 615 VDWI 618

QY 813 VPWI 816

RESULT 3

ENTRY #type complete

TITLE plasma kallikrein (EC 3.4.21.34) precursor - mouse

ORGANISM #formal_name Mus musculus #common_name house mouse

DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change

18-Jun-1999

ACCESSIONS A36557

REFERENCE A36557

#authors Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachepa, L.; Rochemont, J.; Maki, M.; Chretien, M.

#journal DNA Cell Biol. (1990) 9:737-748

#title Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison of protein and mRNA levels


```
391-621      experimental #label MAT2\
21-104,47-77,51-57,      #domain trypsin homology #label TRY\
111-194,137-166,
141-147,201-284,
227-256,231-237,
292-375,318-347,
322-328,340-345,
363-503,419-435,
517-584,548-563,
574-602
127,215,308,453,
459,494      #disulfide_bonds #status predicted\
#binding_site carbohydrate (Asn) (covalent) #status
#predicted\
#binding_site carbohydrate (Asn) (covalent) #status
#binding_site carbohydrate (Asn) (covalent) #status
#active_site His, Asp, Ser #status predicted
#active_site His, Asp, Ser #status predicted
SUMMARY      #length 638 #molecular-weight 71273 #checksum 227
434,483,578      #length 638 #molecular-weight 71273 #checksum 227
#local Match      31.4%; Score 594, DB 1; Length 638;
#local Similarity 38.9%; Pred. No. 2,92e-109;
#conservative 60; Mismatches 71; Indels 18; Gaps 16;
#conservative 60; Mismatches 71; Indels 18; Gaps 16;
Db 391 IYGGTSSSGEMPWQVSLQVLT-SQNHM-CGGSIIGRWITLAHCFDIPDPWRIY 448
OY 578 IIGKNSLGGMPWQVSLKSSHGDRLCGATLLSSCWLTAAHCFK--RIGNSTRSY 635
Db 449 SG-IINLT-SEITNK-TPFSIKELIHOKYKMGSESDIALIKLQPLN-YTEF--OK-P 501
OY 636 AVRVGDYHILVEEPEELGVQOIVIHREYRPSDYDIALVRLQGEEDCARFSSHYLP 695
Db 502 ICLP-SKADFTIYNCGWVTGKSGEKEGTONILQKATIPLVNECCQKRYDYITQ 560
OY 696 ACLPLMRERPKTASNCYITGMDT-GRAYSRF-LQQAAPLPLPKRCEERYKG-RFTGR 752
Db 561 MCAGY-KEGK-IDACKGDSGGLVCKHSG-RWQLVGTSWEGSCARKREPPYITKVAEY 617
OY 752 MCAGNLEHHRKRVDSQSGGLVCKHSGRPEESWVYGVTSWGYCGVDPGYITKVAEY 812
Db 618 IDWI 621
OY 813 VPWI 816
RESULT 5
ENTRY      KOHUP      #type complete
TITLE      plasma kallikrein (EC 3.4.21.34) precursor - human
ALTERNATE_NAMES      kininogen; plasma prekallikrein
ORGANISM      #formal_name Homo sapiens #common_name man
#sequence_revision 13-Aug-1986 #text_change
18-Jun-1999
#cross-references GB:113143; NID:g190262; PIDN:AAA60153.1; PID:g190263
#accession A00921
#molecule_type mRNA
#residues 1-638 #label CHU
#cross-references GB:113143; NID:g190262; PIDN:AAA60153.1; PID:g190263
REFERENCE      A37939
#authors      MCMullen, B.A.; Fujikawa, K.; Davie, E.W.
#journal      Biochemistry (1991) 30:2050-2056
#title      Location of the disulfide bonds in human plasma
#title      prekallikrein: the presence of four novel apple domains in
#title      the amino-terminal portion of the molecule.
#cross-references MUID:91152016
#accession A37939
#molecule_type protein
#residues 20-27,40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;
#residues 103-113;131-140;141-143;'S',144-146;147-159;187-193,
```


Db 617 YMDWI 621
QY 812 FVPMI 816

RESULT 6
ENTRY A34170 #type complete
TITLE acrosin (EC 3.4.21.10) precursor - pig
ALTERNATE_NAMES 53k fucose-binding protein
ORGANISM #formal name Sus scrofa domestica #common name domestic pig
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
ACCESSIONS A34170; S08994; S02428; S04940; S16657; S02780; S10695;
S12968

ENCE A34170
#authors Baba, T.; Kashiwabara, S.; Watanabe, K.; Itoh, H.; Michikawa,
Y.; Kimura, K.; Takada, M.; Fukamizu, A.; Arai, Y.
#journal J. Biol. Chem. (1989) 264:11920-11927
#title Activation and maturation mechanisms of boar acrosin zymogen
based on the deduced primary structure.
#cross-references M01D:89308595
#accession A34170
#status Preliminary
#molecule_type mRNA
#residues 1-415 #label BAB
#cross-references GB:J04950; NID:g164702; PIDN:AAA31131.1; PID:g164703
S08994

REFERENCE S08994
#authors Cechova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
#journal Biol. Chem. Hoppe-Seyler (1990) 371:317-323
#title Is spermthogen a modified proacrosin? Isolation,
purification, and partial characterization of
low-molecular-mass boar proacrosin.
#cross-references M01D:90253655
#accession S08994
#molecule_type protein
#residues 'X',18,'X',20-25,'X',27-32,'X',34-38,'X',40-50 #label
CBO

REFERENCE S02428
#authors Toepfer-Petersen, E.; Henschen, A.
#journal FEBS Lett. (1987) 226:38-42
#title Acrosin shows zona and fucose binding, novel properties for a
serine proteinase.
#cross-references M01D:88083633
#accession S02428
#molecule_type protein
#residues 17-32;40-55 #label TOE
ENCE S04940
#authors Adam, I.M.; Klemm, U.; Maier, W.M.; Hoyer-Fender, S.;
Tsaousidou, S.; Engel, W.
#journal Eur. J. Biochem. (1989) 182:563-568
#title Molecular cloning of preproacrosin and analysis of its
expression pattern in spermatogenesis.
#cross-references M01D:89325301
#accession S04940
#molecule_type mRNA
#residues 1-7,9-210,'Q',212-216,'VT',219-346,'A',348-388,390-393,
'GN',396,'LVE',399-409,'RRTARLLI' #label ADH
#cross-references EMBL:X14844
#note the authors translated the codon CCT for residue 240 as
Ala, GCC for residue 264 as Gly, and ACC for residue
298 as Ser

REFERENCE S16657
#authors Adam, I.M.
#submission submitted to the EMBL Data Library, March 1989
#accession S16657
#molecule_type mRNA
#residues 1-7,9-210,'Q',212-216,'VT',219-346,'A',348-388,390-398,
'KELL', #label AD2
#cross-references EMBL:X14844; NID:g1867; PIDN:CAA32948.1; PID:g1868
#note the difference at the carboxyl end is due to a
frameshift error

REFERENCE S02780

#authors Baba, T.; Michikawa, Y.; Kawakura, K.; Arai, Y.
#journal FEBS Lett. (1989) 244:132-136
#title Activation of boar proacrosin is effected by processing at
both N- and C-terminal portions of the zymogen molecule.
#cross-references M01D:89171246
#accession S02780
#molecule_type protein
#residues 17-69 #label BA2

REFERENCE S10695
#authors Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.;
Zucker, A.
#journal FEBS Lett. (1990) 265:51-54
#title Zona pellucida-binding of boar sperm acrosin is associated
with the N-terminal peptide of the acrosin B-chain (heavy
chain).
#cross-references M01D:90306316
#accession S10695
#molecule_type protein
#residues 40-62 #label TO2

REFERENCE S12968
#authors Toepfer-Petersen, E.; Calvete, J.; Schaefer, W.; Henschen, A.
#journal FEBS Lett. (1990) 275:139-142
#title Complete localization of the disulfide bridges and
glycosylation sites in boar sperm acrosin.
#cross-references M01D:91085546
#accession S12968
#molecule_type protein
#residues 17-29;34-66;68-91;94-121;123-166;171-184;190-207;
209-216;219-228;231-245;248-381;383-415 #label TO3

CLASSIFICATION #superfamily acrosin; trypsin homology
KEYWORDS glycoprotein; hydrolase; serine proteinase; sperm
FEATURE
1-16 #domain signal sequence #status predicted #label SIG\
17-415 #product acrosin #status experimental #label MAT\
17-39 #product acrosin light (A) chain #status experimental
40-415 #product acrosin heavy (B) chain #status experimental
40-283 #domain LCH\
300-374 #domain LCH\
19,208 #domain LCH\
22-152,26-160, #domain LCH\
71-87,175-244, #domain LCH\
207-223,234-264 #domain LCH\
86,140,238 #domain LCH\
SUMMARY #disulfide bonds #status experimental\
#active-site His, Asp, Ser #status predicted
#length 415 #molecular-weight 45387 #checksum 6203

Query Match 30.3%; Score 574; DB 1; Length 415;
Best Local Similarity 39.1%; Pred. No. 1,65e-104;
Matches 99; Conservative 56; Mismatches 79; Indels 19; Gaps 14;

Db 40 VVGKSAEFGKAPFPMWVSLDIFMTHNNRHYTGGLINSHWVITAAHCKRKKKVTDMRL 99
QY 578 IIGKNSLGGWPMQVSLKSKSHGGR-L-LOGATLLSSCWVITAAHCKRKGNS--R- 633
Db 100 IIGANVWVSGKPKVPRLOERF-VEEIIHKKYVGLINLALIKI-TPPPGCPF-- 155
QY 634 STAVR-VG-DYHIVLPEEEFEELGVOQIVIHREYRDRSDYDIALVRLQGPEDCARFSS 691
Db 156 -IGPGCLPQFKAGPPRAPOTCWVYGWYKLEKGRPTSPLOEARVALIDLELNSFRWYN 214
QY 692 HVLPAQLPMBRERPKTANCYITGNG--DGRVSRLLQQAIFLLKRCDE-R-YK 746
Db 215 GRIRSTNVCACTPRG-K-IDTCOGDSGGLPKCRDRAENTFVVGITSGVGCARAKRPV 272
QY 747 GRGTGMLCAGNIIHEKRVDSOCGDSGGLMC-ERPGESWVYGYVTSWGYCGVXDSPV 805
Db 273 YSTWTPYLNWIAS 285
QY 806 YTKVSAFVPMIS 816

RESULT 7
ENTRY A37344 #type complete
TITLE acrosin (EC 3.4.21.10) precursor form 1 - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
ACCESSIONS A37344
REFERENCE A37344
#authors Klemm, U.; Maier, W.M.; Tsousidou, S.; Adham, I.M.;
#journal Willison, K.; Engel, W.
#title Differentiation (1990) 42:160-166
#cross-references CDNA sequence, primary structure and postmeiotic expression in spermatogenesis.
#accession A37344
#status preliminary
#molecule_type mRNA
#residues 1-418 #label KLE
#cross-references GB:X52466; NID:949857; PIDN:CAA36704.1; PID:949858
#classification #superfamily acrosin; trypsin homology
#note glycoprotein; hydrolase; serine proteinase
#domain trypsin homology #label TRY
#binding_site carbohydrate (Asn) (covalent) #status predicted
#disulfide_bonds #status predicted
#disulfide_bonds #status predicted
#disulfide_bonds #status predicted
#active_site His, Asp, Ser #status predicted
#disulfide_bonds #status predicted
#disulfide_bonds #status predicted
#disulfide_bonds #status predicted
#disulfide_bonds #status predicted
#length 418 #molecular_weight 46816 #checksum 9559
SUMMARY
Query Match 29.6%; Score 560; DB 2; Length 418;
Best Local Similarity 40.7%; Pred. No. 3,436-101;
Matches 103; Conservative 48; Mismatches 84; Indels 18; Gaps 13;
Db 39 IVSGGSAHWGAMPWWSLQIFITSHNSRRYHACGSLNSHWLTAHCFDNKKRYDWR 98
Qy 578 IIGKNSLGGWQWVSLRKSLSHGDRLLCGATLLSCWVLAHCF--KRYGSTR- 633
Db 99 VFGAHEIEGRNPKVPEQERYVQKIVHEKYNVTEGNDIALKV-TPPYTCGDF--- 154
Qy 634 SYAVRGDHTLVP-EFEFEIEGVQOIVIHREYRPPRSYDIALVRLOGPEOCARFSH 692
Db 155 IGPCCLPHERKAGPPRIKPHCYVTGWGYSIKREAPRSPVLEAARVDLIDLCNSTQWYN 214
Qy 693 VLPACLPIMRERPOKTAASNCYITGMDTGRAVS-R--LQQAIPLPKRFCE--ERYK 746
Db 215 GAVTSTNVCAGY-PEGR-IDTCGDSGGLPMLCRDNRQPFVYVIGTISWVGACARARPGV 272
Qy 747 GFTGRMLCAGNLHEHRRVDSGCGSGPLMC-ERPGESWVYGVYISWVGCGVKDPSGV 805
Db 273 TATWDYLDWIAS 285
Qy 806 YTKVSAFVPIKS 818
RESULT 8
ENTRY S18407 #type complete
TITLE acrosin (EC 3.4.21.10) precursor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
ACCESSIONS S18407; S30037; A56620
REFERENCE S18407
#authors Klemm, U.; Flake, A.; Engel, W.
#journal Biochim. Biophys. Acta (1991) 1090:270-272
#title Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic origin.
#cross-references MUID:92031708

#accession S18407
#molecule_type mRNA
#residues 1-437 #label KLE
#cross-references EMBL:X59254
REFERENCE S30037
#authors Klemm, U.; Flake, A.; Engel, W.
#submission submitted to the EMBL Data Library, April 1991
#accession S30037
#molecule_type mRNA
#residues 1-254, 'LCDR', 259, 'DHEL', 264, 'GRLC', 269-437 #label KLE2
#cross-references EMBL:X59254; NID:957282; PIDN:CAA1947.1; PID:957283
REFERENCE A56620
#authors Klemm, U.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
#journal DNA Seq. (1991) 2:57-60
#title Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
#cross-references MUID:92199245
#accession A56620
#status preliminary
#molecule_type DNA; mRNA
#residues 1-254, 'LCDR', 259, 'DHEL', 264, 'GRLC', 269-437 #label KRE
#note sequence modified after extraction from NCBI backbone
#sequence extracted from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447, NCBIN:89454, NCBIN:89458)
#superfamily acrosin; trypsin homology
#glycoprotein; hydrolase; serine proteinase; sperm; zymogen
CLASSIFICATION
KEYWORDS 1-19
20-43
43-286
44-437
22, 211
25-155, 29-163
74-90
89, 143, 241
178-247
210-226
237-267
SUMMARY
#length 437 #molecular_weight 48279 #checksum 1220
Query Match 29.4%; Score 557; DB 2; Length 437;
Best Local Similarity 40.9%; Pred. No. 1,766-100;
Matches 103; Conservative 46; Mismatches 86; Indels 17; Gaps 11;
Db 43 IVGGQSSRMAMPWWSLQIFITSHNSRRYHACGSLNSHWLTAHCFDNKKRYDWR 102
Qy 578 IIGKNSLGGWQWVSLRKSLSHGDRLLCGATLLSCWVLAHCF--KRYGSTR- 633
Db 103 VFGAHEIEGRNPKVPEQERYVQKIVHEKYNVTEGNDIALKV-TPPYTCGDF--- 158
Qy 693 VLPACLPIMRERPOKTAASNCYITGMDTGRAVS-R--LQQAIPLPKRFCE--ERYK 747
Db 159 VFGGCLPHERKAGPPRIKPHCYVTGWGYSIKREAPRSPVLEAARVDLIDLCNSTQWYN 218
Qy 748 RFTGRMLCAGNLHEHRRVDSGCGSGPLMC-ERPGESWVYGVYISWVGCGVKDPSGV 806
Db 219 RVNTSTNVCAGY-PEGR-IDTCGDSGGLPMLCRDNRQPFVYVIGTISWVGACARARPGV 276
Qy 748 RFTGRMLCAGNLHEHRRVDSGCGSGPLMC-ERPGESWVYGVYISWVGCGVKDPSGV 806
Db 277 TATWDYLDWIAS 288
Qy 807 YTKVSAFVPIKS 818
RESULT 9
ENTRY S47538 #type complete
TITLE acrosin (EC 3.4.21.10) precursor - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit

```
##Residues      4-436  #Label KAS
##Cross-references GB:D00754; NID:g220322; PIDN:BAA00651.1;
```

02.000220	A31597	REFERENCE
	Ny. E. J. Leonardson G. J. Hensch & T. M. Jantrowski	Authors

```

#journal      DNA (1988) 7:671-677
#title        Cloning and characterization of a cDNA for rat tissue-type
#cross-references M01D:89170114
#accession    A31597
#molecule-type mRNA
#residues     1-379, 'K', 381-559 #label NYT
#cross-references GB:M2697; NID:9530159; PIDN:AAA1812.1; PID:9530160
CLASSIFICATION #superfamily tissue plasminogen activator: EGF homology;
                #fibronectin type I repeat homology; kringle homology;
                #trypsin homology
                #fibrinolysis: glycoprotein: hydrolase: kringle: serine
                #proteinase

FEATURE
1-17          #domain signal sequence #status predicted #label SIG\
18-29         #domain propeptide #status predicted #label PRO\
30-559        #product t-plasminogen activator #status predicted
                #label MAT\
30-308        #product t-plasminogen activator chain A #status
                #predicted #label ACH\
30-75         #domain fibronectin type I repeat homology #label 1F1\
30-116        #domain EGF homology #label EGF\
124-205       #domain kringle homology #label KR1\
213-294       #domain kringle homology #label KR2\
309-559       #product t-plasminogen activator chain B #status
                #predicted #label BCH\
309-553       #domain trypsin homology #label TRY\
38-68,66-75,83-94,
88-105,107-116,
124-205,145-187,
176-200,213-284,
234-276,265-289,
297-428,340-356,
348-417,442-516,
474-490,506-534
149,481       #disulfide_bonds #status predicted\
                #binding_site carbohydrate (asn) (covalent) #status
                #predicted\
308-309       #cleavage_site Arg-Ile (plasmin, trypsin) #status
                #predicted\
355,404,510   #active_site His, Asp, Ser #status predicted
                #length 559 #molecular_weight 62903 #checksum 8573

SUMMARY
Query Match      28.8%; Score 545; DB 1; Length 559;
Best Local Similarity 38.3%; Pred. No. 7.03e-98;
Matches 93; Conservative 58; Mismatches 76; Indels 16; Gaps 13;

Db 321 PQOAFVKKRSPGPEFLGCVLSSCWVSAHCEVERF--PPHHKVLGRTYR-VV 377
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
647 PEEFEIEIGVOQIVIHREYRPRSDYDIALVRLGPEQCARFSHVLPACLPMLREHPQ 706
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
378 PGEFEQTEIEKXIVHKEFDDYTDNDIALQLRSDSSQCAQESSVGTACLPD-PDYOL 436
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
437 PMTECELSGYKHEASSPFSDRLKEAHVLYSSRSCTSOHLNKTITSMNLCAGDRT 496
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
707 KTASNCYITLGMG--DTRRAY-SRTLOQAIPLPKRRC-EER-YKGRPTGMLCAGNLH- 760
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
497 GGNQDVHADACGSGGGLVC-MIDKRMTLGIIISMGLGCGKDVPGITTKTYNLINMTQD 555
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
761 -EKRRV-DSGCGSGGLPKMERGESWVYGVTSWGTGCKVDSGVYTKVASAFVPIWIKS 818
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
556 NMK 558
819 VTK 821

RESULT 12
ENTRY      S29599      #type fragment
TITLE      acrosin (EC 3.4.21.10) precursor - guinea pig (fragment)
ORGANISM   #formal name Cavia porcellus - common name guinea pig
DATE       22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
           22-Jun-1999

```

```

ACCESSIONS  S29599
REFERENCE    S29599
#authors    Gerton, G.L.; Hoff, H.B.; Baba, T.
#submitted  submitted to the EMBL Data Library, May 1992
#description The amino acid sequence of guinea pig proacrosin deduced from
              its cDNA sequence.
#accession  S29599
#molecule-type mRNA
#residues   1-421 #label GER
#cross-references EMBL:212153; NID:949559; PIDN:CAA78137.1; PID:949560
CLASSIFICATION #superfamily acrosin: trypsin homology
KEYWORDS      glycoprotein: hydrolase: serine proteinase
FEATURE
41-264        #domain trypsin homology #label TRY
SUMMARY
Query Match      28.8%; Score 545; DB 2; Length 421;
Best Local Similarity 39.1%; Pred. No. 1.21e-97;
Matches 99; Conservative 46; Mismatches 89; Indels 19; Gaps 14;

Db 41 IIGQTAOPGAMPVMSLQIFMAHNNRRYHACGILLNSHWVLTAAHCEFSKRYDWRL 100
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
578 IIGKNSLRGMPQVSLRKSXSHGDR-LCGATLLSCWVLTAAHCF--KRYGSTR- 633
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
101 VFGAELEIYGNKKRYRPLQERY-VEKIVTHEKNIVNEGNDALALKITPPVS-CGPF-- 156
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
634 SYAVRVGDY--HTLVPEEIEIGVOQIVIHREYRPRSDYDIALVRLGPEQCARFS 691
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
157 -IGPGCLPTFRAGPKIPQICVAVAGVYIREKAPRPSVLEARELIDDLGNSQWYN 215
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
692 HVLACLPMLREPKQKTAISNCTYITGMDT-GRV-Y-SRTLOQAIPLPKRFE--ERYK 746
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
747 GRFGRLMCAQNLHHRKRVDSGCGSLMC-ERPGESWVYGVTSWGTGCKVDSGPV 805
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
274 YTAWTDYLDWIAS 286
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
806 YTKVASAFVPIWIKS 818

RESULT 13
ENTRY      UKR1
TITLE      u-plasminogen activator (EC 3.4.21.73) precursor - human
ALTERNATE_NAMES cellular plasminogen activator; urokinase-type
                  plasminogen activator (uPA)
CONTAINS    urokinase-type plasminogen activator chain A; urokinase-type
              plasminogen activator chain A1; urokinase-type plasminogen
              activator chain B; urokinase-type plasminogen activator,
              single chain form
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        17-Dec-1982 #sequence_revision 04-Dec-1986 #text_change
           18-Jun-1999
ACCESSIONS  A00931; I52209; J0102; A37561; I38102; S65783; A37562;
              A37563; A37564; A35689; A36697
REFERENCE    A00931
#authors    Ricciio, A.; Grimaldi, G.; Verde, P.; Sebastio, G.; Boast, S.;
              Biasi, F.
#journal     Nucleic Acids Res. (1985) 13:2759-2771
#title       The human urokinase-plasminogen activator gene and its
              promoter.
#cross-references M01D:85215647
#accession  A00931
#molecule-type DNA
#residues   1-431 #label RIC
#cross-references GB:X02419; NID:937601; PIDN:CAA26268.1; PID:e300604;
           PTD:91834524
           the authors translated the codon ATG for residue 214 as
           Ile
REFERENCE    I52209
#authors    Nagamine, Y.; Pearson, D.; Gratian, M.
#journal     Biochem. Biophys. Res. Commun. (1985) 132:563-569
#title       Exon-Intron boundary sliding in the generation of two mRNAs

```

coding for porcine urokinase-like plasminogen activator.

#cross-references MUID:86050639

#accession 152209

#status preliminary; translated from GB/EMBL/DBJ

#molecule-type DNA

#residues 145-161 ##label NAG1

#cross-references GB:K03027; NID:g340174; PIDN:AAA61257.1; PID:g340175

REFERENCE

#authors Nagai, M.; Hiramatsu, R.; Kaneda, T.; Hayasuke, N.; Arimura, H.; Nishida, M.; Suyama, T.

#journal Gene (1985) 36:183-188

#title Molecular cloning of cDNA coding for human preprourokinase.

#cross-references MUID:86056554

#accession J10102

#molecule-type mRNA

#residues 1-213,'T',215-431 ##label NAG2

#cross-references GB:K03226; NID:g340155; PIDN:AAC97138.1; PID:g340158; GB:D00244; NID:g220138; PID:d1000623; PID:g220139

REFERENCE

#authors A37561

#journal Verde, P.; Stoppelli, M.P.; Galeffi, P.; Di Nocera, P.; Blasi, F.

#title Proc. Natl. Acad. Sci. U.S.A. (1984) 81:4727-4731

#cross-references GB:K03226; NID:g220138; PID:d1000623; PID:g220139

#accession A37561

#molecule-type mRNA

#residues 66-431 ##label YER

#cross-references GB:D00244; NID:g220138

REFERENCE

#authors 138102

#journal Jacobs, P.; Cravador, A.; Loriau, R.; Brockly, F.; Colau, B.; Chuchana, P.; van Elsen, A.; Herzog, A.; Bollen, A.

#title DNA (1985) 4:139-146

#title Molecular cloning, sequencing, and expression in *Escherichia coli* of human preprourokinase cDNA.

#cross-references MUID:85203359

#accession 138102

#status preliminary

#molecule-type mRNA

#residues 1-150,'W',152-213,'T',215-385,'C',387-429,'V',431

#label JAC

#cross-references EMBL:X02760; NID:g35297; PIDN:CAA26535.1; PID:g35298

REFERENCE

#authors 565783

#journal Yoshimoto, M.; Ushiyama, Y.; Sakai, M.; Tamaki, S.; Hara, H.; Takahashi, K.; Sawasaki, Y.; Hanada, K.

#title Biochim. Biophys. Acta (1996) 1293:83-89

#title Characterization of single chain urokinase-type plasminogen activator with a novel amino-acid substitution in the kringle structure.

#cross-references MUID:96186279

#accession S65783

#status preliminary

#molecule-type mRNA

#residues 21-140,'L',142-213,'T',215-431 ##label YOS

#cross-references EMBL:D11143; NID:g1311467; PIDN:BAA01919.1; PID:d1002396; PID:g1199928

REFERENCE

#authors A37562

#journal Gunzler, W.A.; Steffens, G.J.; Otting, F.; Kim, S.M.A.; Franks, E.; Flohe, L.

#title Hoppe-Seyler's Z. Physiol. Chem. (1982) 363:1155-1165

#title The primary structure of high molecular mass urokinase from human urine.

#cross-references MUID:83055084

#accession A37562

#molecule-type protein

#residues 21-177 ##label GUN

REFERENCE

#authors A37563

#journal Schaller, J.; Nick, H.; Rickli, E.E.; Gillesse, D.; Lergier, W.; Studer, R.O.

#title Eur. J. Biochem. (1982) 125:251-257

#title Human low-molecular-weight urinary urokinase. Partial characterization and preliminary sequence data of the two polypeptide chains.

#cross-references MUID:83003608

#accession A37563

#molecule-type protein

#residues 156-176;179-193,'T',195,'T',197-224 ##label SCH

REFERENCE

#authors A37564

#journal Steffens, G.J.; Gunzler, W.A.; Otting, F.; Franks, E.; Flohe, L.

#title Hoppe-Seyler's Z. Physiol. Chem. (1982) 363:1043-1058

#title The complete amino acid sequence of low molecular mass urokinase from human urine.

#cross-references MUID:83055099

#accession A37564

#molecule-type protein

#residues 158-410 ##label STE

REFERENCE

#authors A35689

#journal Kenter, E.J.; Buko, A.; Menon, G.; Sarin, V.K.

#title Biochem. Biophys. Res. Commun. (1990) 171:401-406

#title Carbohydrate composition and presence of a fucose-protein linkage in recombinant human pro-urokinase.

#cross-references MUID:90365737

#accession A35689

#molecule-type protein

#residues 21-30,'X',32,'X',34-38,'X',40-43 ##label KEN

#note Identification of a fucose and attempt to determine its attachment site

REFERENCE

#authors A36697

#journal Rabbani, S.A.; Desjardins, J.; Bell, A.W.; Banville, D.; Mazur, A.; Henkin, J.; Goltzman, D.

#title Biochem. Biophys. Res. Commun. (1990) 173:1058-1064

#title An amino-terminal fragment of urokinase isolated from a prostate cancer cell line (PC-3) is mitogenic for osteoblast-like cells.

#cross-references MUID:91097529

#accession A36697

#molecule-type protein

#residues 21-34 ##label RAB

REFERENCE

#authors A51255

#journal Li, X.; Bokman, A.M.; Llinas, M.; Smith, R.A.G.; Dobson, C.M.

#title Submission submitted to the Brookhaven Protein Data Bank, July 1993

#cross-references PDB:1KDU

#contents annotation; conformation and disulfide bond assignments by (1)H-NMR, residues 69-153

REFERENCE

#authors A44375

#journal Li, X.; Smith, R.A.G.; Dobson, C.M.

#title Biochemistry (1992) 31:9562-9571

#title Sequential (1)H NMR assignments and secondary structure of the kringle domain from urokinase.

#cross-references MUID:93003110

#contents annotation; conformation and disulfide bond assignments by (1)H-NMR

REFERENCE

#authors A66822

#journal Hansen, A.P.; Petros, A.M.; Meadows, R.P.; Nettlesham, D.G.; Mazur, A.P.; Olejniczak, E.T.; Xu, R.X.; Pederson, T.M.; Henkin, J.; Resik, S.W.

#title Submission submitted to the Brookhaven Protein Data Bank, January 1994

#cross-references PDB:1U9K

#contents annotation; conformation and disulfide bond assignments by (1)H-NMR, residues 26-155

REFERENCE

#authors A66058

#journal Spraggon, G.S.; Phillips, C.; Nowak, U.K.; Ponting, C.P.; Saunders, D.; Dobson, C.M.; Stuart, D.I.; Jones, E.Y.

#title Submission submitted to the Brookhaven Protein Data Bank, July 1995

#cross-references PDB:1LWV

#contents annotation; X-ray crystallography, 2.5 angstroms, residues 168-175;179-426

COMMENT

This enzyme is found in urine in a high molecular mass form, consisting of A and B chains, and a low molecular mass form, consisting of A1 and B chains.

COMMENT

Urokinase-type plasminogen activator proteolytically activates plasminogen, and the inactive single-chain form is proteolytically activated by plasmin (see PR:PLHU).

GENETICS

#gene GDB:PLAU


```

#accession      JE0315      heart.
#status          preliminary
#molecule_type mRNA
#residues        1-1113 #label TOM
#cross-references DBJ:AB013874
CLASSIFICATION  #superfamily trypsin homology
FEATURE
869-1097        #domain trypsin homology #label TRY
SUMMARY          #length 1113 #molecular-weight 122984 #checksum 224

Query Match      28.6%; Score 542; DB 2; Length 1113;
Best Local Similarity 40.7%; Pred. No. 6.19e-97;
Matches 98; Conservative 39; Mismatches 90; Indels 14; Gaps 11;

869  ILGRTSRPGKMPWQCSIQ--SEPS-GHI-CGCVLIARKWVLTVAHCFEGREDAD-VWKV 923
      |::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
578  ILGKNSLRGMPWQVSLRLKSSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSTRSYAV 637
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 924  VEGINNLDPHPSGFEMQTRFVKITILLHPYRSRAVVDYDISVEL--SDI-NE-TSYVRPVC 979
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 638  RVGDYHTLVPEEFEEIGVOQIVIHREYRPDRSDYDALVRLQPEEQCAFSSSHVLPAC 697
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 980  LP-SPEEYLEBDITYCYITGMGHMGNKMPKIQEGEVRIIPLEQCQSYFDMKTTNRMICA 1038
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 698  LPLMREPRORTASNCYITIGWDTGRAYSRILQCAIPLPRFCERYKGR-FTGRMLCA 756
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1039  G--YESGTVDSMGDSGGPLVCERPGGQWTLFGLTSMGVCFSKVLGPVYSNYSYFVGW 1096
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 757  GNLHEKRVDSQCQDSGGLMCEKRGESWVYGVTSWYGCGVKD-SPGVYTKVSAFVPM 815
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1097  I 1097
      |
QY 816  I 816

```

Search completed: Mon Mar 13 10:18:43 2000
 Job time : 20 secs.

THIS PAGE BLANK (USPTO)

Asse mit No. 1

(a) (b) (c) (d) (e)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

    -pp protein - protein database search, using Smith-Waterman algorithm
Run on:      Mon Mar 13 10:29:12 2000;      MasPar time 4.72 Seconds
Tabular output not generated.      277.462 Million cell updates/sec

```

```

Title: >US-09-147-947-6
Description: (227-327) from US09147947A.pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLGGSVHGEGRNELYHAG.....KSMGEHNCKGHEKDAQVSCCT 101

```

Scoring table: PAM 150

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database: (a-issued)

Statistics: Mean 26.593; Variance 104.340; scale 0.255

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

ID	Score	Query Length	DB	ID	Description	Pred. No.
1	435	57.5	4	5510466-4	Patent No. 5510466.	1.34e-34e
2	428	56.5	3	PCT-US86-0	Sequence 2, Applicatio	7.26e-34e
3	428	56.5	4	US-08-973-	Sequence 2, Applicatio	7.26e-34e
4	428	56.5	1	US-08-953-	Sequence 2, Applicatio	7.26e-34e
5	428	56.5	4	US-08-948-	Sequence 2, Applicatio	7.26e-34e
6	427	56.4	1	US-08-154-	Sequence 2, Applicatio	9.24e-34e
7	412	54.4	5	US-08-673-	Sequence 10, Applicati	3.44e-33e
8	412	54.4	2	US-08-916-	Sequence 10, Applicati	3.44e-33e
9	412	54.4	3	US-08-977-	Sequence 10, Applicati	3.44e-33e
10	365	48.2	1	US-08-470-	Sequence 10, Applicati	2.75e-21e
11	355	46.9	2	US-08-994-	Sequence 7, Applicatio	3.00e-28e
12	355	46.9	5	US-08-392-	Sequence 2, Applicatio	1.72e-24e
13	338	44.6	4	US-08-994-	Sequence 2, Applicatio	1.72e-24e
14	338	44.6	2	US-08-994-	Sequence 2, Applicatio	1.72e-24e
15	125	16.5	1	US-08-200-	Sequence 6, Applicatio	2.35e-03e
16	125	16.5	3	PCT-US84-0	Sequence 2, Applicatio	2.35e-03e
17	88	11.6	1	US-08-681-	Sequence 1, Applicatio	4.20e+00e
18	83	11.0	1	US-08-477-	Sequence 2, Applicatio	1.09e+01e
19	83	11.0	1	US-08-881-	Sequence 2, Applicatio	1.09e+01e
20	83	11.0	1	US-08-176-	Sequence 2, Applicatio	1.09e+01e
21	83	11.0	1	US-08-179-	Sequence 2, Applicatio	1.09e+01e
22	83	11.0	1	US-08-100-	Sequence 2, Applicatio	1.09e+01e
23	79	10.4	1	US-08-291-	Sequence 4, Applicatio	2.31e+01e

45	76	10.0	1178	1	US-08-446-	Sequence 5, Applicatio	4.03e+01
44	76	10.0	1178	1	US-08-446-	Sequence 5, Applicatio	4.03e+01
43	76	10.0	1178	1	US-08-446-	Sequence 5, Applicatio	4.03e+01
42	76	10.0	1177	1	US-07-920-	Sequence 8, Applicatio	4.03e+01
41	76	10.0	1177	1	US-07-920-	Sequence 8, Applicatio	4.03e+01
40	76	10.0	1177	1	US-07-920-	Sequence 8, Applicatio	4.03e+01
39	76	10.0	1176	3	PCT-US93-1	Sequence 10, Applicati	4.03e+01
38	76	10.0	1176	3	PCT-US93-1	Sequence 10, Applicati	4.03e+01
37	76	10.0	1165	1	PCT-US92-1	Sequence 14, Applicati	4.03e+01
36	76	10.0	1165	1	PCT-US92-1	Sequence 14, Applicati	4.03e+01
35	76	10.0	1165	1	US-08-356-	Sequence 4, Applicatio	4.03e+01
34	76	10.0	1165	1	US-08-356-	Sequence 4, Applicatio	4.03e+01
33	76	10.0	425	1	US-08-190-	Sequence 58, Applicatio	4.03e+01
32	76	10.0	425	1	US-08-190-	Sequence 58, Applicatio	4.03e+01
31	76	10.0	23	1	US-08-477-	Sequence 11, Applicati	4.03e+01
30	77	10.2	3111	2	US-08-477-	Sequence 11, Applicati	4.03e+01
29	77	10.2	3111	2	US-08-477-	Sequence 11, Applicati	4.03e+01
28	77	10.2	3111	2	US-08-477-	Sequence 11, Applicati	4.03e+01
27	77	10.2	1130	2	US-08-135-	Sequence 2, Applicatio	3.35e+01
26	77	10.2	1130	2	US-08-135-	Sequence 2, Applicatio	3.35e+01
25	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
24	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
23	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
22	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
21	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
20	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
19	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
18	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
17	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
16	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
15	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
14	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
13	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
12	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
11	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
10	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
9	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
8	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
7	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
6	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
5	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
4	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
3	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
2	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01
1	79	10.4	1168	3	US-08-460-	Sequence 2, Applicatio	3.35e+01

ALIGNMENTS

RESULT	1		STANDARD;	PRT;	491 AA.
ID	5510466-4				
XX	xxxxxx				
XX	01-JAN-1900				
DT					
XX	Patent No. 5510466.				
DE					
XX					
CC	Patent No. 5510466				
CC	APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO				
CC	TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY				
CC	THERETO				
CC	NUMBER OF SEQUENCES: 12				
CC	CURRENT APPLICATION DATA:				
CC	APPLICATION NUMBER: US/08/307,400				
CC	FILING DATE: 16-SEP-1994				
CC	PRIOR APPLICATION DATA:				
CC	APPLICATION NUMBER: 997,113				
CC	FILING DATE: 24-DEC-1992				
CC	APPLICATION NUMBER: 391,486				
CC	FILING DATE: 09-AUG-1989				
CC	APPLICATION NUMBER: 272,002				
CC	FILING DATE: 15-NOV-1988				
CC	SEQ ID NO:4:				
CC	LENGTH: 453				
CC	SEQUENCE 491 AA; 54317 MW; 1350575 CN;				
QO					
	Query Match	57.5%;	Score 435;	DB 4;	Length 453;
	Best Local Similarity	54.5%;	Pred. No. 1,34e-34;		
	Matches 55;	Conservative 18;	Mismatches 28;	Indels 0;	Gaps 0
Db	352 VRLVGGSGPHBGRVYEIFHEGQGWTCDDRWELRGGLVVCRSLSIGYKGVSVHKKRAYFGKGT 411				
OY	227 IRLAGGSVHEBGRVELYHAGQGWTCDDQMDADAEVLCRQLGSLGIAKAMHQAIFYGEGS 286				
Db	412 GPIMLNEVYCFKPRESSIFECCRIRQMGVYRACGSHDDAGYTCT 452				
OY	287 GPVMLDEYRCGTGNELSLDQCPKSSWGEHNCGHKEDAGVSCVT 327				
RESULT	2		STANDARD;	PRT;	451 AA.
ID	PCT-US96-08081-2				
XX	xxxxxx				
XX					
XX					
DT					


```
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jervis, Herbert H.
CC REGISTRATION NUMBER: 31,171
CC REFERENCE/DOCKET NUMBER: SBC-P50338
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (610) 270-5019
CC TELEFAX: (610) 270-5090
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SS SEQUENCE 451 AA; 49762 MW; 1024670 CN;
Dy ery Match 56.5%; Score 428; DB 1; Length 451;
Dy st Local Similarity 53.5%; Pred.No.7,26e-34;
Dy Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0
Db 350 VRLVGGSGPHGGRVEILHSGWGITICDDRWEVRVGQVYCRSLGYPGVQAHHKAAHFQGT 409
Qy 227 IRLAGGSVSHHGREGVELYNHAGQMGTCDQDDQMDADAEVLCIGLSTAKAMHQAYFEGS 286
Qy 410 GPIMLNVEYFCFGRESSIECKIRONGTFRACSHSDAGYTCT 450
Qy 287 GPVMIDEVRCGNELSIEOCRKSSWGEHNCGHKEDAGVSC 327
XX RESULT 5 STANDARD; PRT; 451 AA.
XX ID US-08-948-222-2
XX AC xxxxxx
XX DT
XX DE Sequence 2, Application US/08948222
XX CC
XX CC Sequence 2, Application US/08948222
XX CC Patent No. 5863798
XX CC GENERAL INFORMATION:
XX CC APPLICANT: Lytko, Paul G.
XX CC APPLICANT: Elshourbagy, Nabil A.
XX CC APPLICANT: Brawner, Mary E.
XX CC TITLE OF INVENTION: Attachment Enhanced 293 Cells
XX CC NUMBER OF SEQUENCES: 4
XX CC CORRESPONDENCE ADDRESS:
XX CC ADDRESSEE: SmithKline Beecham - Corporate Patents
XX CC ADDRESSEE: U.S.
XX CC STREET: Mailcode - UW2220, 709 Swedeland Road
XX CC CITY: King of Prussia
XX CC STATE: Pennsylvania
XX CC COUNTRY: U.S.A.
XX CC ZIP: 19406-5090
XX CC COMPUTER READABLE FORM:
XX CC MEDIUM TYPE: Floppy disk
XX CC COMPUTER: IBM PC compatible
XX CC OPERATING SYSTEM: PC-DOS/MS-DOS
XX CC SOFTWARE: PatentIn Release #1.0, Version #1.30
XX CC CURRENT APPLICATION DATA:
XX CC APPLICATION NUMBER: US/08/948,222
XX CC FILING DATE:
XX CC CLASSIFICATION:
XX CC PRIOR APPLICATION DATA:
XX CC APPLICATION NUMBER: US/08/453,117
XX CC FILING DATE:
XX CC ATTORNEY/AGENT INFORMATION:
XX CC NAME: Jervis, Herbert H.
XX CC REGISTRATION NUMBER: 31,171
XX CC REFERENCE/DOCKET NUMBER: SBC-P50338
XX CC TELECOMMUNICATION INFORMATION:
XX CC TELEPHONE: (610) 270-5019
XX CC TELEFAX: (610) 270-5090
```

```

CC INFORMATION FOR SEQ ID NO: 2
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 451 AA; 49762 MM; 1024670 CN;
SQ
Query Match 56.5%; Score 428; DB 2; Length 451;
Best Local Similarity 53.5%; Pred. No. 7.26e-34;
Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
Db 350 VRLVGGSGPHRGREIILHSGGWMGICDDREVRNGVGVVCSTLPGVQAHHKAHFGGT 409
Oy 227 IRLGGSSVHGQVELHAGVGVCDDQWDADAIEYICRQLGSLGAKMHOAYFEEGS 286
Db 410 GPIWLNEVFCFGRESSIEECKIRQMGTRACSHSEDAVGTC 450
Oy 287 GPVMLDEYRCGNELSLIEGCTKSSWGHNCGHKEDAGVSC 327
RESULT 6 STANDARD: PRT; 451 AA.
ID US-08-154-365-2
XX xxxxxx
DT
XX
DE
XX
Sequence 2, Application US/08154365
CC Sequence 2, Application US/08154365
CC Patent No. 5624904
CC GENERAL INFORMATION:
CC APPLICANT: Dunne, Dana W.
CC APPLICANT: Resnick, David
CC APPLICANT: Kreiger, Monty
CC APPLICANT: Joiner, Keith A.
CC TITLE OF INVENTION: Method for Treating Gram-Positive
CC TITLE OF INVENTION: Septicemia
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Patrea L. Pabst
CC STREET: 1100 Peachtree Street, Suite 2800
CC CITY: Atlanta
CC STATE: Ga
CC COUNTRY: USA
CC ZIP: 30309-4530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/154,365
CC FILING DATE:
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: MIT6392
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404)-815-6508
CC TELEFAX: (404)-815-6555
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: YES
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal

```

CC ORIGINAL SOURCE:
CC ORGANISM: homo sapien
CC PUBLICATION INFORMATION:
CC AUTHORS: Ashkenas, et al.
CC JOURNAL: J. Lipid Res.
CC VOLUME: 34
CC PAGES: 983-1000
CC DATE: 1993
CC RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 451
SQ SEQUENCE 451 AA: 49841 MW: 1025052 CN:

Query Match 56.4%; Score 427; DB 1; Length 451;
Best Local Similarity 53.5%; Pred. No. 9,24e-34;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

Db 350 VRLVGGSPHGRVETLLSGGWTICDDNWEVGVGVCRSLGYPGVAVHRAHFGQCT 409
227 IRLAGSSVHEGRVELYHAGWGTVCDQDMDADAEVICRDLGSLAKAWHQAYFEGSS 286
410 GPIMLNEVFCFGRESSIECKIRQWGTACSHSEDAVYCT 450
QY 287 GPVIMDEVCTGNELSIEQCPKSSWGEHNCGHKEDAGVCT 327

RESULT 7
ID US-08-473-791-10 STANDARD; PRT; 585 AA.
XX
XX
AC xxxxxx
DT
DE
XX
XX

Sequence 10, Application US/08473791
Patent No. 5736340
GENERAL INFORMATION:
CC APPLICANT: Kolhs, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casipit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/473.791
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316.714
CC FILING DATE:
CC APPLICATION NUMBER: US/07/961.404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 585 AA: 65330 MW: 1874129 CN:

Query Match 54.4%; Score 412; DB 1; Length 585;
Best Local Similarity 53.5%; Pred. No. 3.44e-32;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

Db 24 MRLAGGATNGRVIETFRGQWGYVDNLMDLTDASVCRALGFENATQALGRAFGQGS 83
QY 227 IRLAGSSVHEGRVELYHAGWGTVCDQDMDADAEVICRDLGSLAKAWHQAYFEGSS 286
Db 84 GPIMLDEVCTGTEASLADCKSLGWLKSNCRHERDAGVCT 124
QY 287 GPVIMDEVCTGNELSIEQCPKSSWGEHNCGHKEDAGVCT 327

RESULT 8
ID US-08-316-714-10 STANDARD; PRT; 585 AA.
XX
XX
AC xxxxxx
DT
DE
XX
XX

Sequence 10, Application US/08316714
Patent No. 5965382
GENERAL INFORMATION:
CC APPLICANT: Kolhs, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casipit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316.714
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/961.404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear

CC MOLECULE TYPE: Protein
SQ SEQUENCE 585 AA; 65330 MW; 1874129 CN;
Query Match 54.4%; Score 412; DB 2; Length 585;
Best Local Similarity 53.5%; Pred. No. 3,44e-32;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;
DB 24 MRLADGAINOGVEVEFYRGOWGTVCNDLMDLTDASVCRALGFENATQALGRAAFGGGS 83
OY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQMDADAEVICRQLGSLGIAKAMHQAIFYEGGS 286
DB 84 GPMLDEVCTGTAEASLADCKSLGWLKSNCRHERDAGVCT 124
OY 287 GPVMLDEVCTGTAEASLADCKSLGWLKSNCRHERDAGVCT 327
T 9
US-08-477-674-10 STANDARD; PRT: 585 AA.
AC xxxxxx
DE Sequence 10, Application US/08477674
XX
XX
XX
CC Sequence 10, Application US/08477674
CC Patent No. 5644035
CC GENERAL INFORMATION:
CC APPLICANT: Kolhs, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casibit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/477,674
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE:
CC APPLICATION NUMBER: US/07/961,404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2395.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA; 65330 MW; 1874129 CN;

Query Match 54.4%; Score 412; DB 1; Length 585;
Best Local Similarity 53.5%; Pred. No. 3,44e-32;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;
DB 24 MRLADGAINOGVEVEFYRGOWGTVCNDLMDLTDASVCRALGFENATQALGRAAFGGGS 83
OY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQMDADAEVICRQLGSLGIAKAMHQAIFYEGGS 286
DB 84 GPMLDEVCTGTAEASLADCKSLGWLKSNCRHERDAGVCT 124
OY 287 GPVMLDEVCTGTAEASLADCKSLGWLKSNCRHERDAGVCT 327
RESULT 10
ID US-08-470-350B-2 STANDARD; PRT: 1290 AA.
XX
XX
XX
AC xxxxxx
DE Sequence 2, Application US/08470350B
XX
XX
XX
CC Sequence 2, Application US/08470350B
CC Patent No. 5684126
CC GENERAL INFORMATION:
CC APPLICANT: Li, Xiao
CC APPLICANT: Snyder, Solomon H.
CC TITLE OF INVENTION: Edoetin: A Secreted von Ebner's Gland
CC TITLE OF INVENTION: Protein Associated with Taste Buds
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1001 G Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,350B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wolfe, Susan A.
CC REGISTRATION NUMBER: 33,568
CC REFERENCE/DOCKET NUMBER: 01107,48790
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-508-9100
CC TELEFAX: 202-508-9299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1290 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1290 AA; 141879 MW; 9315616 CN;
Query Match 48.2%; Score 365; DB 1; Length 1290;
Best Local Similarity 52.4%; Pred. No. 2,75e-27;
Matches 54; Conservative 17; Mismatches 28; Indels 4; Gaps 4;
DB 292 VRLVNGDRGRCGRVETLQ-GSMGTVCDSWDTKDANYVCRQL-VCGWALSAPGSAHFQ 349
OY 227 IRLAGSSVHEGRVE-LYHAGOMGTVCDDQMDADAEVICRQLGSLG-IKAMHQAIFYEG 284
DB 350 GSGSTVLDVACTGHEAYLMSCSHRGWLSHNCGHEDAGVICS 392
OY 285 GSGPVMLDEVCTGTAEASLADCKSLGWLKSNCRHERDAGVCT 327

RESULT 11
ID US-08-794-795-7 STANDARD; PRT: 489 AA.
AC xxxxxx
XX
XX
DT
XX
Sequence 7, Application US/08794795
XX Sequence 7, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Eishourilagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Smithline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 489 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 489 AA; 49441 MW; 1141296 CN;
SO
Query Match 46.9%; Score 355; DB 2; Length 489;
Best Local Similarity 50.5%; Prid. No. 3.00e-26;
Matches 51; Conservative 23; Mismatches 22; Indels 5; Gaps 5;
Db 394 VRIMGSTN-R-GRAEYNNENGTICDDDDNNDAIVFCMLGYS-RGRA-LSSY-GGGS 448
OY 227 IRLAGSSVHEGRVELYHAGQGTVCDDWDADAIVICRQLGSLGIAKAMHQAIFEGGS 286
DB 449 GNIMLDNVNCRGTENSLMDCSKNSMGNNHCNVHEDAGVECS 489
OY 287 GPVMDDEVCTGNETLSIEQPKSSWGEHNCGRKEDAGVCT 327
RESULT 12
ID US-08-392-367B-2 STANDARD; PRT: 518 AA.
AC xxxxxx
XX
XX
DT
XX

XX
DT
XX
DE
XX
XX
Sequence 2, Application US/08392367B
CC Patent No. 5691197
CC GENERAL INFORMATION:
CC APPLICANT: Tryggvason, Karl
CC APPLICANT: Eliomaa, Outi
CC APPLICANT: Kangas, Maarit
CC TITLE OF INVENTION: An Insolated DNA Sequence For a
CC Patent No. 5691197
CC TITLE OF INVENTION: No. 5691197a1 Macrophage Receptor with
CC TITLE OF INVENTION: a Collagenous Domain and the
CC TITLE OF INVENTION: Polypeptide Chain Encoded by
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fay, Sharpe, Beall, Fagan,
CC STREET: Minnich & McKee
CC STREET: 1100 Superior Avenue
CC CITY: Cleveland
CC STATE: Ohio
CC COUNTRY: U.S.A.
CC ZIP: 44114-2518
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch,
CC MEDIUM TYPE: 720 Kb storable
CC COMPUTER: IBM PS/2, Model 35 SX
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: Word Perfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/392,367B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Minnich, Richard J.
CC REGISTRATION NUMBER: 24,175
CC REFERENCE/DOCKET NUMBER: TRY 2 009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (216) 861-5582
CC TELEFAX: (216) 241-1666
CC TELEX: (216) 980162
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 518 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: Single
CC TOPOLOGY: Linear
CC SEQUENCE 518 AA; 52730 MW; 1281000 CN;
SO
Query Match 46.9%; Score 355; DB 1; Length 518;
Best Local Similarity 50.5%; Prid. No. 3.00e-26;
Matches 51; Conservative 23; Mismatches 22; Indels 5; Gaps 5;
Db 423 VRIMGSTN-R-GRAEYNNENGTICDDDDNNDAIVFCMLGYS-RGRA-LSSY-GGGS 477
OY 227 IRLAGSSVHEGRVELYHAGQGTVCDDWDADAIVICRQLGSLGIAKAMHQAIFEGGS 286
DB 478 GNIMLDNVNCRGTENSLMDCSKNSMGNNHCNVHEDAGVECS 518
OY 287 GPVMDDEVCTGNETLSIEQPKSSWGEHNCGRKEDAGVCT 327
RESULT 13
ID US-08-794-795-2 STANDARD; PRT: 495 AA.
AC xxxxxx
XX
XX
DT
XX

DE Sequence 2, Application US/08794795
XX
CC Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eplor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA: 49764 MW: 1218428 CN:
CC
CC Query Match 44.6%; Score 338; DB 2; Length 495;
CC Best Local Similarity 46.5%; Pred. No. 1,72e-24;
CC Matches 47; Conservative 25; Mismatches 24; Indels 5; Gaps 4;
DB 399 VRIVGSSN-R-GRAEYVYSGTGTICDDEWQNSDAIVFCMIGYS-KGRALYKV--GAGT 453
QY 227 IRLAGSSVHEGVEYLHAGWGTVCDWDADAIVICROLGSLAKAMHQAIFYEGGS 286
DB 454 GOIWLNVOCRGTESTLWSTCKNSMGHDCSHEDAGVCS 494
QY 287 GPVMDVEVRCTGNELSLIEQCPKSSMGHNGHREDAGVSC 327
RESULT 14
ID US-08-794-795-6 STANDARD: PRT: 520 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08794795
XX
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:

CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eplor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA: 52658 MW: 1340662 CN:
CC
CC Query Match 44.6%; Score 338; DB 2; Length 520;
CC Best Local Similarity 46.5%; Pred. No. 1,72e-24;
CC Matches 47; Conservative 25; Mismatches 24; Indels 5; Gaps 4;
DB 424 VRIVGSSN-R-GRAEYVYSGTGTICDDEWQNSDAIVFCMIGYS-KGRALYKV--GAGT 478
QY 227 IRLAGSSVHEGVEYLHAGWGTVCDWDADAIVICROLGSLAKAMHQAIFYEGGS 286
DB 479 GOIWLNVOCRGTESTLWSTCKNSMGHDCSHEDAGVCS 519
QY 287 GPVMDVEVRCTGNELSLIEQCPKSSMGHNGHREDAGVSC 327
RESULT 15
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08200900A
XX
CC Sequence 2, Application US/08200900A
CC Patent No. 5665566
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

(W.T.)

Distribution rights by Oxford Molecular Ltd

1 IRLAGGSSVHEGRVELYHAG.....KSSWGEHNCGHKEDAGVCT 101

2AM 150
3ap 11

42080 seqs, 47172406 residues

existing first 45

pir62
1:pir1 2:pir2 3:pir3 4:pir4

mean 37.625; Variance 64.653; scale 0.582

and is derived by analysis of the total score distribution.

SUMMARIES

Accession	Protein	Description	Pred. No.
U55719	CS5759	brain-specific serine	1,86-138
U55719	CS5759	LDL receptor I, macro	6,47e-77
U55719	CS5759	macrophage scavenger	4,66e-75
U55719	CS5759	macrophage scavenger	1,15e-73
U55719	CS5759	macrophage scavenger	3,33e-73
U55719	CS5759	macrophage scavenger	3,33e-73
U55719	CS5759	scavenger receptor cy	5,68e-73
U55719	CS5759	M30 antigen - human	1,99e-70
U55719	CS5759	M30 antigen (extracel	1,99e-70
U55719	CS5759	M30 antigen (cytosol	1,99e-70
U55719	CS5759	M30 antigen (cytosol	1,99e-70
U55719	CS5759	Mac-2-binding glycop	1,67e-69
U55719	CS5759	macin (clone pGM1-1)	5,73e-67
U55719	CS5759	scavenger receptor cy	4,77e-66
U55719	CS5759	cyclophilin C-cassioa	4,61e-63
U55719	CS5759	macin (clone pGM-1)	2,24e-62
U55719	CS5759	antigen WC1.1 precurs	1,52e-60
U55719	CS5759	ebhrin precursor - r	1,02e-58
U55719	CS5759	macrophage bacteriab	1,92e-56
U55719	CS5759	T-cell glycoprotein C	1,17e-48
U55719	CS5759	meed6 precursor - mou	9,25e-45
U55719	CS5759	spectact receptor pro	1,67e-42
U55719	CS5759	enteropeptidase (EC 3	6,11e-14

... in the

45	83	11.0	1680	2	A43434	
44	83	11.0	494	2	T05302	
43	84	11.1	2825	2	T14271	
42	84	11.1	741	2	A47063	
41	84	11.1	558	2	T15448	
40	85	11.1	76	2	T39960	
39	85	11.2	78	2	S76817	
38	85	11.4	1113	2	TE0315	low-density lipoprote
37	86	11.4	810	2	D67050	probable membrane pro
36	87	11.5	226	2	D9156	conserved hypothetica
35	90	11.9	527	2	B70920	hypothetical protein
34	91	12.0	603	2	S76615	hypothetical protein
33	95	12.5	425	2	B64316	restriction modifical
32	99	13.1	494	1	A29079	lymphocyte surface gl
31	101	13.3	501	1	A43356	T-cell surface glycop
29	103	13.6	403	2	T02513	complement C3b/C4b i
28	104	13.7	613	2	S15468	hypothetical protein
27	112	14.8	495	1	A26396	T-cell surface glycop
26	125	16.5	1035	1	A43090	gene CD5 protein - sh
25	126	16.6	1019	1	A56318	entriopeptidase (EC 3
24	127	16.8	495	1	16611370	T-cell surface glycop

ALIGNMENTS

RESULT	1	
ENTRY	JC5759	#type complete
TITLE	brain-specific serine proteinase (EC 3.4.21.-) - mouse	
ORGANISM	#formal_name Mus musculus #common_name house mouse	
DATE	24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 17-Mar-1999	
ACCESSIONS	JC5759	
REFERENCE	JC5759	
authors	Yamamura, Y.; Yamashiro, K.; Tsuruoka, N.; Nakazato, H.; Tsujimura, A.; Yamaguchi, N.	
#journal	Biochem. Biophys. Res. Commun. (1997) 233:386-392	
#title	Molecular cloning of a novel brain-specific serine protease with a kringle-like structure and three scavenger receptor cysteine-rich motifs.	
#cross-references	NCBI:98008848	
#accession	JC5759	
#molecule_type	mRNA	
#residues	1-761	#label YAM
#cross-references	DBSU:DB9871	
#experimental_source	brain	
CLASSIFICATION	#superfamily trypsin homology; scavenger receptor cysteine-rich domain homology	
KEYWORDS	glycoprotein; hydrolase; serine proteinase	
FEATURE	85-157	#domain kringle-like #status predicted #label KRI\
	163-266	#domain scavenger receptor cysteine-rich domain homology #label SRC\
	166-266,273-372,	
	386-486	#domain scavenger receptor cysteine-rich #status predicted #label SRC\
	513-516	#domain furin binding #status predicted #label FRB\
	517-755	#domain trypsin homology #label TRY\
	93,521,569	#binding_site carbohydrate (asn) (covalent) #status predicted\
	562,612,711	#active_site His, Asp, Ser #status predicted
SUMMARY	#length 761	#molecular_weight 84136 #checksum 5449
Query Match	93.0%;	Score 704; DB 2; Length 761;
Best Local	Similarity 90.0%;	Pred. No. 1,86c-188;
Matches	90; Conservative 7;	Mismatches 13; Indels 0; Gaps 0;
Db	166	IRLVGSGSGEGVEVYHAGQMGTCDDQWDADADVCROLGSGTAKAHQAHPFGS 225
		: : : : : : : : :
OY	227	IRLAGGSSVHEGRVETVYHAGQMGTCDDQWDADADVEYICROLGSGTAKAHQAHPFGS 286
		: : : : : : : :
Db	226	GPILDEVRCGTGEMSIEDCPKSSWGEHNCGHKEPDAGVSC 265
		: : : : : : : :

...

Query Match 56.4%; Score 427; DB 2; Length 2153;

Db	462	VARVLGLNNEGVVEFLNNQKGTVCDDDMGTDPANVYVCGLTPSCGSARSSAYFERGS	521
Qy	227	IRLAGGSSVHEGVVELYHMAQGMGTVCDDQMDADAEVYICRLGSLGIAKAMHQAIFEGGS	286
Db	522	VPILIDNAGCGSNGERSELDICSNNGIGVHNCGHODEAVVCT	562
Qy	287	GPVMDDEVKCTGNELSLIEQCPKSSWGEHNCGHKEDAGVST	327

RESULT	8
ENTRY	S36077
TITLE	M130 antigen human
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Aug-1996
IONS	138003; S36077
NCE	138003
thors	Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
#journal	Eur. J. Immunol. (1993) 23:2320-2325
#title	A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
#cross-references	MUID:93380506
#accession	138003
#status	preliminary: translated from GB/EMBL/DBJ
##molecule_type	mRNA
##residues	1-1116 #label RES
##cross-references	EMBL:223968; NID:9312141; PID:9312142
CLASSIFICATION	#superfamily scavenger receptor cysteine-rich domain homology
FEATURE	43-147
-151-254	#domain scavenger receptor cysteine-rich domain homology #label SRC1\
258-361	#domain scavenger receptor cysteine-rich domain homology #label SRC2\
365-468	#domain scavenger receptor cysteine-rich domain homology #label SRC3\
470-573	#domain scavenger receptor cysteine-rich domain homology #label SRC4\
575-678	#domain scavenger receptor cysteine-rich domain homology #label SRC5\
711-814	#domain scavenger receptor cysteine-rich domain homology #label SRC6\
816-920	#domain scavenger receptor cysteine-rich domain homology #label SRC7\
-1024	#domain scavenger receptor cysteine-rich domain homology #label SRC8\
Y	#domain scavenger receptor cysteine-rich domain homology #label SRC9
Query Match	55.0%: Score 416; DB 2; Length 1116;
Best Local Similarity	55.4%: Pred. No. 1.99e-70;
Matches	56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

Db	924	IRLQGPSTSGRVEITWGGSMGTVCDDSDMDLDAQVCCQLCGPALKFAKEAFSGQT	983
Qy	227	IRLAGGSSVHEGVVELYHMAQGMGTVCDDQMDADAEVYICRLGSLGIAKAMHQAIFEGGS	286
Db	964	GPVIMLVNVCCKGNESLMDCPARRMGHSECGHKEDAIVNCT	1024
Qy	287	GPVMDDEVKCTGNELSLIEQCPKSSWGEHNCGHKEDAGVST	327

RESULT	9
TITLE	138006
ORGANISM	M130 antigen (extracellular variant) - human
DATE	17-May-1996 #sequence_revision 17-May-1996 #text_change 02-Aug-1996
ACCESSIONS	138006; S35768

```

REFERENCE
#authors      Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
#journal      Eur. J. Immunol. (1993) 23:2320-2325
#title        A new macrophage differentiation antigen which is a member of
               the scavenger receptor superfamily.
#cross-references MIMD:93380506
#accession    I38006
               #status      preliminary; translated from GB/EMBL/DBJ
               #molecule_type mRNA
               #residues     1-1149 #label RES
CLASSIFICATION #cross-references EMBL:Z22971; NID:9312147; PID:9312148
               #superfamily scavenger receptor cysteine-rich domain homology
FEATURE
43-147        #domain scavenger receptor cysteine-rich domain homology
               #label SRC1\
151-254       #domain scavenger receptor cysteine-rich domain homology
               #label SRC2\
258-361       #domain scavenger receptor cysteine-rich domain homology
               #label SRC3\
365-468       #domain scavenger receptor cysteine-rich domain homology
               #label SRC4\
470-573       #domain scavenger receptor cysteine-rich domain homology
               #label SRC5\
608-711       #domain scavenger receptor cysteine-rich domain homology
               #label SRC6\
744-847       #domain scavenger receptor cysteine-rich domain homology
               #label SRC7\
849-953       #domain scavenger receptor cysteine-rich domain homology
               #label SRC8\
954-1057      #domain scavenger receptor cysteine-rich domain homology
               #label SRC9
SUMMARY
               #length 1149 #molecular-weight 124328 #checksum 487
Query Match      55.0%; Score 416; DB 2; Length 1149;
Best Local Similarity 55.4%; Pred. No. 1,99e-70;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

Db 957 IRLQEGTSGSGRAYEIHGGSGWGTCCDDSDMDLDAQVQCOQLGCGPALAKFAEFGGT 1016
    11 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
Qy 227 IRLAGSSVHEGRRELYHAGQMGTCVDDQDDPADEVICROLGLGIAKAWHQAAYFGGS 286
    11 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

Db 1017 GPVWLNEVKCKGNSSLMDCPARKWHSCGKHEDAAVACT 1057
    11 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
Qy 287 GPVWLDEVRCTGNELSTEQCPKSSWGHNCGRKEDAGVCT 327
    11 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

RESULT 10
ENTRY      I38004      #type complete
TITLE      M130 antigen (cytosolic variant 1) - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       17-May-1996 #sequence_revision 17-May-1996 #text_change
              07-Feb-1997
ACCESSIONS I38004; I38005; I38006; I38007; I38008
REFERENCE
#authors    Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
            Willis, A.C.; Mason, D.Y.
#journal    Eur. J. Immunol. (1993) 23:2320-2325
#title      A new macrophage differentiation antigen which is a member of
            the scavenger receptor superfamily.
#cross-references MIMD:93380506
#accession   I38004
            #status      preliminary; translated from GB/EMBL/DBJ
            #molecule_type mRNA
            #residues     1-1151 #label RES
CLASSIFICATION #cross-references EMBL:Z22966; NID:9312143; PID:9312144
            #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS     cytosol
              #domain scavenger receptor cysteine-rich domain homology
              #label SRC1\
151-254      #domain scavenger receptor cysteine-rich domain homology
              #label SRC2\

```



```
##residues      1-577 ##label CHI
##cross-references EMBL:X67809, NID:g297032, PID:g297033, GB:X67803
##note          sequence appears consistent with either a GPI anchor or
```

REFERENCE A48231 transmembrane domain near the carboxyl end

#authors Friedman, J.; Trahey, M.; Weissman, I.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6815-6819
#title Cloning and characterization of cyclophilin C-associated protein: a candidate natural cellular ligand for cyclophilin C.

#cross-references M01D:9342080

#accession A48231

#molecule-type mRNA
#residues 1-24, 'G', 26-227, 'P', 229-465, 'NE', 468-572, 'LH' ##label FRI

GENETICS ##cross-references GB:L16894; NID:9397799; PID:9397800

CLASSIFICATION CycAP
#gene #superfamily scavenger receptor cysteine-rich domain homology
#domains cell surface component; glycoprotein; membrane protein
#label SRC
#domain signal sequence #status predicted #label STG
#domain scavenger receptor cysteine-rich domain homology
#label SRC

SUMMARY #length 577 #molecular-weight 64491 #checksum 6506

Query Match 50.7%; Score 384; DB 2; Length 577;
Best Local Similarity 50.5%; Pred. No. 4,61e-63;
Matches 51; Conservative 15; Mismatches 35; Indels 0; Gaps 0;

Db 24 MRLVNGASANEGRVEIFRGRMGTVCDNLWNLDAHVCRALGYENATOLGRAAFGPGK 83

QY 227 IRLAGSSVHEGRVELYHAGOWGVCDQWDADAEYICROLGLSGTAKAMHQAIFEGGS 286

Db 84 GPIMLDEVCTGTSSLASCRSLGMYVSRGHEKDAVCS 124

QY 287 GPVWLDEVCTGNELSIQCPSKSWGEHNCGRKDAVCSCT 327

Search completed: Mon Mar 13 10:28:30 2000
Job time: 13 secs.

THIS PAGE BLANK (USPTO)

(2分)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

h_pp protein - protein database search, using Smith-Waterman algorithm
      Mon Mar 13 10:27:15 2000;      MasPar time 6.46 Seconds
      466.945 Million cell updates/sec
Tabular output not generated.

```

```

Title: >US-09-147-947-6
Description: (227-327) from US09147947A.ppt (4 of 6)
Perfect Score: 757
Sequence: 1 IRLGGSSVHEGRPELYTHAG.....KSSMEHNHCGRKEDAGVCT 101

```

Scoring table: PAM 150

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38

Statistics: Mean 38.6669; Variance 61.361; scale 0.6300

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	757	100.0	875	1	NETP_HUMAN	NEUTROTRYPsin PRECURSOR	2,71e-133
2	708	93.5	761	1	NETM_MOUSE	NEUTROTRYPsin PRECURSOR	1.46e-150
3	444	58.7	453	1	MSRE_BOVIN	MACROPHAGE SCAVENGER R	4,98e-83
4	436	57.6	454	1	MSRE_RABIT	MACROPHAGE SCAVENGER R	5,04e-81
5	430	56.8	458	1	MSRE_MOUSE	MACROPHAGE SCAVENGER R	1,60e-79
6	428	56.5	451	1	MSRE_HUMAN	MACROPHAGE SCAVENGER R	5,07e-79
7	373	49.3	1436	1	MC1I_BOVIN	ANTIGEN MC1.1	2,33e-65
8	325	42.9	468	1	CD6_HUMAN	T-CELL DIFFERENTIATION	1,30e-55
9	293	38.7	552	1	SPER_STRPU	EGG PEPTIDE SPERACT RE	6,69e-46
10	158	20.9	1034	1	ENTR_PIG	ENTEROPEPTIDASE PRECUR	3,64e-15
11	127	16.8	495	1	CD5_BOVIN	T-CELL SURFACE GLYCOPR	7,27e-09
12	126	16.6	1019	1	ENTR_HUMAN	ENTEROPEPTIDASE PRECUR	1,14e-08
13	125	16.5	1035	1	ENTR_BOVIN	ENTEROPEPTIDASE PRECUR	1,78e-08
14	120	15.9	1069	1	ENTR_MOUSE	ENTEROPEPTIDASE (EC 3.	1,63e-07
15	112	14.8	495	1	CD5_HUMAN	T-CELL SURFACE GLYCOPR	5,18e-06
16	101	13.3	451	1	CD5_RAT	T-CELL SURFACE GLYCOPR	4,99e-04
17	99	13.1	494	1	CD5_MOUSE	T-CELL SURFACE GLYCOPR	1,12e-03
18	95	12.5	343	1	Y13O_METJA	HYPOTHETICAL PROTEIN M	5,42e-03
19	92	12.2	492	1	TMS2_HUMAN	TRANSMEMBRANE PROTEASE	1,73e-02
20	91	12.0	603	1	SYT_SYNY3	THROMBIN-TRNA SYNTHETA	2,53e-02
21	84	11.1	741	1	GUN5_CLOTH	ENDOGUCANASE SS PRECU	3,39e-01
22	83	11.0	1660	1	FURIN_DROME	FURIN-LIKE PROTEASE	4,86e-01
23	81	10.0	212	1	YM95_MCTCU	HYPOTHETICAL 23.3 KD P	9,88e-01

RESULT	1	STANDARD:	PRF:	875 AA.
ID	NEUR_HUMAN			
AC	P56730:			
DT	15-DEC-1999 (Rel. 39, Creasead)			
DT	15-DEC-1999 (Rel. 39, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	NEUROTRYPsin PRECURSOR (EC 3.4.21.-) (MOTOPsin).			
GN	PRSS12.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN:			
RX	MEDLINE: 98201705.			
RA	PROBA K., GSCHWEND T.P., SONDERGGER P.:			
RT	"Cloning and sequencing of the cDNA encoding human neurotrypsin.";			
RL	Biochim. Biophys. Acta 1396:143-147(1998).			
CC	-1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 4 SRCR DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).			
CC	-----			
DR	EMBL: AJ001531; CAA04816.1; -			
DR	PROSITE: PS00134; TRYPsin_HIS. 1.			
DR	PROSITE: PS00135; TRYPsin_SER. 1			
DR	PROSITE: PS00420; SPERACT_RECEPTOR. 3.			
DR	Hydrolase: Serine protease; GLYCOProtein; Kringle; Repeat; Signal.			
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	875	NEUROTRYPsin.
FT	DOMAIN	23	92	PROLINE-RICH.
FT	DOMAIN	93	165	KRINGLE.
FT	DOMAIN	170	271	SRCR 1.
FT	DOMAIN	280	381	SRCR 2.
FT	DOMAIN	387	487	SRCR 3.

```

FT DOMAIN 500 601 SRCR 4
FT DOMAIN 619 875 SERINE PROTEASE.
FT DOMAIN 619 875 ZYMOGEN ACTIVATION REGION.
FT ACT_SITE 630 630 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 67523272B CRC32:

Query Match 100.0%: Score 757; DB 1; Length 875;
Best Local Similarity 100.0%: Pred. No. 2,71e-163;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 280 IRLAGSSVHEGRVELYHAGOMGTICDDQMDADAEVICHROLGSLGAKAMHQAIFGEGS 339
227 IRLAGSSVHEGRVELYHAGOMGTICDDQMDADAEVICHROLGSLGAKAMHQAIFGEGS 286
340 GPVMLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 380
287 GPVMLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 327

RESULT 2
ID NEUR_MOUSE STANDARD; PRT: 761 AA.
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEURORIPSPIN PRECURSOR (EC 3.4.21.-) (MOTOPSPIN) (BRAIN-SPECIFIC SERINE
DE PROTEASE 3) (BSSP-3).
GN PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GSCHWEND T.P., KRUGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neurotyspin, a novel multidomain serine protease expressed in the
RT nervous system.";
RL Mol. Cell. Neurosci. 9:207-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSURUOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
RT kringle-like structure and three scavenger receptor cysteine-rich
RT motifs.";
RL Biochem. Biophys. Res. Commun. 239:386-392(1997).
RN [1]
RP FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
RP ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
RP LEARNING AND MEMORY OPERATIONS.
RN [1]
RP TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
RN [1]
RP AND AMYGDALA.
RN [1]
RP SUBCELLULAR LOCATION: SECRETED
RN [1]
RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
RN [1]
RP TRYPsin FAMILY.
RN [1]
RP SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
RN [1]
RP SIMILARITY: CONTAINS 3 SRCR DOMAINS.
RN [1]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN [1]
RP between the Swiss Institute of Bioinformatics and the EMBL Outstation -
RN [1]
RP the European Bioinformatics Institute. There are no restrictions on its
RN [1]
RP use by non-profit institutions as long as its content is in no way
RN [1]
RP modified and this statement is not removed. Usage by and for commercial
RN [1]
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN [1]
RP or send an email to license@isb-sib.ch).
RN [1]

```

```

DR EMBL: Y13192; CA72646.1; -
DR EMBL: D89871; BA23986.1; -
DR MGD: MGI:1100881; PRSS12.
DR PFAM: PF00530; SRCR 3.
DR PFAM: PF00089; TRYPsin.1.
DR PROSITE: PS00134; TRYPsin.HIS.1.
DR PROSITE: PS00135; TRYPsin.SER.1.
DR PROSITE: PS00420; SPERACTIN RECEPTOR; 3.
DR Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; signal.
FT SIGNAL 1 21
FT CHAIN 22 761 NEURORIPSPIN.
FT DOMAIN 85 157 POTENTIAL.
FT DOMAIN 166 267 KRINGLE.
FT DOMAIN 273 373 SRCR 1.
FT DOMAIN 386 487 SRCR 2.
FT DOMAIN 505 761 SRCR 3.
FT DOMAIN 505 516 SERINE PROTEASE.
FT ACT_SITE 516 517 ZYMOGEN ACTIVATION REGION.
FT ACT_SITE 562 562 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 CHARGE RELAY SYSTEM.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA: 84118 MW: 3F3CAF35 CRC32:

Query Match 93.5%: Score 708; DB 1; Length 761;
Best Local Similarity 91.0%: Pred. No. 1,46e-150;
Matches 91; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 166 IRLVGSNGHGRVELYHAGOMGTICDDQMDADAEVICHROLGSLGAKAMHQAIFGEGS 225
227 IRLAGSSVHEGRVELYHAGOMGTICDDQMDADAEVICHROLGSLGAKAMHQAIFGEGS 286
226 GPVMLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 265
287 GPVMLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 326

RESULT 3
ID MSRE_BOVIN STANDARD; PRT: 453 AA.
AC P21758;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
DE LDL RECEPTOR I AND II).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RN [1]
RP TISSUE-LUNG;
RX MEDLINE: 90136965.
RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
RA KRIEGER M.;
RT "Type I macrophage scavenger receptor contains alpha-helical and
RT collagen-like coiled coils.";
RL Nature 343:531-535(1990).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RN [1]
RP TISSUE-LUNG;
RX MEDLINE: 90136973.
RA ROHRER L., FREEMAN M., KODAMA T., PENNAN M., KRIEGER M.;
RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
RT scavenger receptor type II.";
RL Nature 343:570-572(1990).
RN [1]
RP FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
RN [1]
RP DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
RN [1]
RP TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIANE THE
RN [1]
RP ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
RN [1]

```



```

DR PFAM: PF01391; Collagen: 1.
KW Transmembrane: Glycoprotein; Endocytosis; Coiled coil; LDL;
KW Heptad repeat pattern; Receptor; Alternative splicing.
KW TRANSMEM 1 55 CYTOPLASMIC (POTENTIAL).
FT FT TRANSMEM 56 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT FT DOMAIN 79 458 (POTENTIAL).
FT FT DOMAIN 79 458 EXTRACELLULAR (POTENTIAL).
FT FT DOMAIN 114 276 SPACER (PROBABLE).
FT FT DOMAIN 277 350 ALPHA-HELICAL COILED-COIL.
FT FT DOMAIN 351 458 COLLAGEN-LIKE.
FT FT CARBOHYD 94 94 CYS-RICH.
FT FT CARBOHYD 107 107 POTENTIAL.
FT FT CARBOHYD 147 147 POTENTIAL.
FT FT CARBOHYD 188 188 POTENTIAL.
FT FT CARBOHYD 253 253 POTENTIAL.
FT FT CARBOHYD 271 271 POTENTIAL.
FT FT VARSPLIC 352 354 TPL -> RSV (IN ISOFORM II).
FT FT VARSPLIC 355 458 MISSING (IN ISOFORM II).
SQ SEQUENCE 458 AA; 50130 MW; F28A556E CRC32;

Query Match 56.8%; Score 430; DB 1; Length 458;
Best Local Similarity 53.5%; Pred. No. 1,606-79;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

Db 357 VRLVGSAGAHGRVDFIFHOGMGITICDDRMDIRAGQVVCRLSGTQEVLAHKRAHFGQGT 416
QY 227 IRLAGSSVHGRVRELYHAGQVTCDDQWDADAEVTCRLDGLSGIAKAWHQAIFYEGS 286
Db 417 GPWLENYWCFGRSSIECNKINQWGVLSGSHSDAGYTC 457
QY 287 GPWMLDEVRCIGNELSTIEQCRKSSNGEHNCGHKEDAGVST 327

RESULT
ID MSRE HUMAN STANDARD; PRT; 451 AA.
AC P21757; P21759;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
DE LDL RECEPTOR I AND II).
GN MSRL.
OS Homo sapiens (Human).
OC Eukaryota; Metacoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
[1]
RN 1 SEQUENCE FROM N.A.
RX MEDLINE; 91067661.
RA MATSUMOTO A., NAITO M., ITAKURA H., IKEMOTO S., ASAKA H.,
RA HAYAKAWA I., KANAMORI H., ABURATANI H., TAKAKU F., SUZUKI H.,
RA KOBAYASHI Y., MIYAI T., TAKAHASHI K., COHEN E.H., WYDRO R.,
RA HOUSMAN D.E., KODAMA T.;
RT "Human macrophage scavenger receptors: primary structure, expression,
RT and localization in atherosclerotic lesions."
RL Proc. Natl. Acad. Sci. U.S.A. 87:9133-9137(1990).
CC -1- FUNCTION: MEMBRANE GLYCOPROTEIN IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
CC OF THE SAME GENE.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```


CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC
CC EMBL: U09860; AAC50138.1; -
CC HSSP: P00763; IDPO.
CC MIM: 226200; -
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00060; MAM_2; 1.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 2.
DR PFAM: PF00057; 1d1_recept_a; 2.
DR PFAM: PF00089; trypsin; 1.
DR PFAM: PF00431; CUB; 2.
DR PFAM: PF00530; SRCR; 1.
DR PFAM: PF00629; MAM; 1.
DR PFAM: PF01390; SEA; 1.
DR Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
DR Serine protease; Zymogen; Transmembrane; Repeat.
KW CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).
KW TRANSMEM 19 47 CATALYTIC CHAIN (LIGHT CHAIN).
KW DOMAIN 182 223 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
KW DOMAIN 225 334 LDL-RECEPTOR CLASS A 1.
KW DOMAIN 342 504 CUB.
KW DOMAIN 524 634 MAM.
KW DOMAIN 641 679 LDL-RECEPTOR CLASS A 2.
KW DOMAIN 678 771 SRCR.
KW ACT_SITE 825 835 CHARGE RELAY SYSTEM (BY SIMILARITY).
KW ACT_SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).
KW ACT_SITE 971 971 CHARGE RELAY SYSTEM (BY SIMILARITY).
KW LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 184 197 BY SIMILARITY.
FT DISULFID 191 210 BY SIMILARITY.
FT DISULFID 204 221 BY SIMILARITY.
FT DISULFID 643 655 BY SIMILARITY.
FT DISULFID 650 668 BY SIMILARITY.
FT DISULFID 662 677 BY SIMILARITY.
FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
FT DISULFID 810 826 BY SIMILARITY.
FT DISULFID 910 977 BY SIMILARITY.
FT DISULFID 941 956 BY SIMILARITY.
FT DISULFID 967 985 BY SIMILARITY.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.

FT CARBOHYD 179 179 POTENTIAL.
FT CARBOHYD 328 328 POTENTIAL.
FT CARBOHYD 335 335 POTENTIAL.
FT CARBOHYD 388 388 POTENTIAL.
FT CARBOHYD 440 440 POTENTIAL.
FT CARBOHYD 470 470 POTENTIAL.
FT CARBOHYD 503 503 POTENTIAL.
FT CARBOHYD 534 534 POTENTIAL.
FT CARBOHYD 630 630 POTENTIAL.
FT CARBOHYD 682 682 POTENTIAL.
FT CARBOHYD 706 706 POTENTIAL.
FT CARBOHYD 725 725 POTENTIAL.
FT CARBOHYD 848 848 POTENTIAL.
FT CARBOHYD 867 867 POTENTIAL.
FT CARBOHYD 909 909 POTENTIAL.
FT CARBOHYD 949 949 POTENTIAL.
SQ SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;
Query Match 16.68; Score 126; DB 1; Length 1019;
Best Local Similarity 27.18; Pred. No. 1,14e-08;
Matches 26; Conservative 23; Mismatches 41; Indels 6; Gaps 6;
Db 678 VREFNGTNNNGLYVRRIOSIMHTACAEWMTQISNDVCGILGL-GSGNSKPIFSDG- 735
QY 227 IRLAGSSVHEGRVEIYHAGQCTVDDQWDADAEVCRQLSLGIAKAMHQAIFGSS 286
Db 736 GPEVKINTAP-DGHLITPSQCLDPSLIRLOCNHR 770
QY 287 GP-VMLDEVACITGN-ELIS-IEQPKSSWGEHNGHK 319
RESULT 13
ID ENTK_BOVIN STANDARD; PRT; 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-DUODENUM;
RX MEDLINE: 94329561.
RA KITAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 94043122.
RA LAVALLE E.R., REHEMUTULA A., RACIE L.A., DIBLASIO E.A.,
RA FERENZ C., GRANT K.L., LIGHT A., MCCOY J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE-INTESTINE;
RX MEDLINE: 92189715.
RA LIGHT A., JANSKA H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN

CC TRYPsinogen.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPsin-LIKE PROTEASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U09859; ABA0026.1; -;
 CC EMBL: L19663; AAA16035.1; -;
 CC PIR: A61436; A61436.
 CC HSP: P00763; IDPO.
 CC PROSITE: PS00134; TRYPsin_HIS. 1.
 CC PROSITE: PS00135; TRYPsin_SER. 1.
 CC PROSITE: PS01180; CUB. 2.
 CC PROSITE: PS00740; MAM_1; 1.
 CC PROSITE: PS50060; MAM_2; 1.
 CC PROSITE: PS01209; LDLR_1; 2.
 CC PROSITE: PS50068; LDLR_2; 2.
 CC PFAM: PF00057; Idl_recept.a; 2.
 CC PFAM: PF00089; trypsin; 1.
 CC PFAM: PF00431; CUB. 2.
 CC PFAM: PF00530; SRCR. 1.
 CC PFAM: PF00629; MAM. 1.
 CC PFAM: PF01390; SEA. 1.
 CC Signal-anchor: Glycoprotein; Myristate; Hydrophobic;
 CC Setine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
 CC CHAIN 1 800
 CC CHAIN 2 1035
 CC TRANSMEM 19 47
 CC DOMAIN 197 238
 CC DOMAIN 240 350
 CC DOMAIN 358 520
 CC DOMAIN 540 650
 CC DOMAIN 657 695
 CC DOMAIN 694 787
 CC ACT_SITE 841 841
 CC ACT_SITE 892 892
 CC ACT_SITE 987 987
 CC LIPID 2 2
 CC DISULFID 199 212
 CC DISULFID 206 225
 CC DISULFID 219 236
 CC DISULFID 659 671
 CC DISULFID 666 684
 CC DISULFID 678 693
 CC DISULFID 788 812
 CC DISULFID 826 842
 CC DISULFID 926 933
 CC DISULFID 957 972
 CC DISULFID 983 1011
 CC CARBOHYD 116 116
 CC CARBOHYD 147 147
 CC CARBOHYD 170 170
 CC CARBOHYD 194 194
 CC CARBOHYD 233 233
 CC CARBOHYD 263 263
 CC CARBOHYD 264 264
 CC CARBOHYD 404 404

FT CARBOHYD 456 456 POTENTIAL.
 FT CARBOHYD 486 486 POTENTIAL.
 FT CARBOHYD 519 519 POTENTIAL.
 FT CARBOHYD 550 550 POTENTIAL.
 FT CARBOHYD 646 646 POTENTIAL.
 FT CARBOHYD 698 698 POTENTIAL.
 FT CARBOHYD 722 722 POTENTIAL.
 FT CARBOHYD 741 741 POTENTIAL.
 FT CARBOHYD 762 762 POTENTIAL.
 FT CARBOHYD 864 864 POTENTIAL.
 FT CARBOHYD 903 903 POTENTIAL.
 FT CARBOHYD 965 965 POTENTIAL.
 FT CARBOHYD 166 192 MISSING (IN SHORT ISOFORM).
 FT VARSPLIC 192
 FT CONFLICT 808
 FT SEQUENCE 1035 AA; 114887 MW; 5173034A CRC32;
 SQ
 Query Match 16.5%; Score 125; DB 1; Length 1035;
 Best Local Similarity 26.8%; Pred. No. 1,78e-08;
 Matches 22; Conservative 21; Mismatches 35; Indels 4; Gaps 4;
 Db 694 VRLNGTDSGLVQFRIOSIMHACAEWNTTQISDDVCLLGL-GTGNSVPT-FSTGG 751
 QY 227 IRLAGSSVHGRVELHACQMGVCCDDADAEVTCRLSLGAKAHQAYFEGRS 286
 Db 752 GPVYNTAPNGSLILPPSQOC 773
 QY 287 GP-VMLDEVRCCTGELS-IEQC 306
 RESULT 14
 ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=DUDENUM;
 RA YUAN X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.;
 Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPsin AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPsinOGEN TO TRYPsin WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPsinOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPsinOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPsin-LIKE PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL: U73378; AAB37317.1; -
 DR HSSP: P00763; IDPO.
 DR MGD; MG1:1197523; PRSS7.
 DR PROSITE: PS00134; TRYPSIN_HIS. 1.
 DR PROSITE: PS00135; TRYPSIN_SER. 1.
 DR PROSITE: PS01180; CUB. 2.
 DR PROSITE: PS00740; MAM. 1; 1.
 DR PROSITE: PS50060; MAM. 2; 1.
 DR PROSITE: PS01209; LDLR. 1; 2.
 DR PROSITE: PS50068; LDLR. 2; 2.
 DR PFAM: PF00057; 1dl_recept.a; 2.
 DR PFAM: PF00089; trypsin. 1.
 DR PFAM: PF00431; CUB. 2.
 DR PFAM: PF00350; SRCR. 1.
 DR PFAM: PF00629; MAM. 1.
 DR PFAM: PF01390; SEA. 1.
 DR Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 Signal-anchor: Glycoprotein; Transmembrane; Repeat;
 CHAIN 1 828
 CHAIN 1 1069
 TRANSMEM 19 47
 DOMAIN 227 268
 DOMAIN 270 379
 DOMAIN 387 549
 DOMAIN 569 679
 DOMAIN 686 724
 DOMAIN 723 816
 DOMAIN 874 874
 ACT_SITE 925 925
 ACT_SITE 1021 1021
 LIPID 2 2
 DISULFID 229 242
 DISULFID 236 255
 DISULFID 249 266
 DISULFID 688 700
 DISULFID 695 713
 DISULFID 707 722
 DISULFID 817 945
 DISULFID 859 875
 DISULFID 959 1027
 DISULFID 991 1006
 DISULFID 1017 1045
 CARBOHYD 147 147
 CARBOHYD 197 197
 CARBOHYD 212 212
 CARBOHYD 373 373
 CARBOHYD 380 380
 CARBOHYD 433 433
 CARBOHYD 515 515
 CARBOHYD 579 579
 CARBOHYD 675 675
 CARBOHYD 727 727
 CARBOHYD 751 751
 CARBOHYD 770 770
 CARBOHYD 791 791
 CARBOHYD 897 897
 CARBOHYD 936 936
 CARBOHYD 999 999
 SEQUENCE 1069 AA; 118735 MW; 48B825A2 CRC32;
 Query Match 15.98; Score 120; DB 1; Length 1069;
 Best Local Similarity 25.38; Pred. No. 1.63e-07;
 Matches 24; Conservative 24; Mismatches 43; Indels 4; Gaps 4;
 Db 723 VRLNGRSNGIYQEMIHSHIACAEENMTQISNEVCHLGL-GSANS-SMPISNGG 780
 QY 227 IRLAGSSVHGRVELYHAQGMGTVCDDQMDADAELVLCROGLSGIAKAWHQAIFGEGS 286
 Db 781 GPEVRVNAQPNGLILTPSLQCSODSLILQCNHR 815
 QY 287 GP-VMLDEVRCGTGNELSIE-QCPKSSWGEHNGCHK 319

RESULT 15
 ID CDS_HUMAN STANDARD; PRT; 495 AA.
 AC P06127;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL SURFACE GLYCOPROTEIN CDS PRECURSOR (LYMPHOCYTE GLYCOPROTEIN
 DE T1/LEU-1) (LYMPHOCYTE ANTIGEN CDS).
 GN CDS OR LEU1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87014786.
 RA JONES N.H., CLABBY M.L., DIALYNAS D.P., HUANG H.-J.S.,
 RA HERZENBERG L.A., SPROMINGER J.L.;
 RT "Isolation of complementary DNA clones encoding the human lymphocyte
 RT glycoprotein T1/Leu-1";
 RL Nature 323:346-349(1986).
 RN [2]
 RP INTERACTION WITH CD72/Lyb-2.
 RX MEDLINE; 91270374.
 RA VAN DE VELDE H., VON HOEGEN I., LHO W., PARNES J.R., THIELEMAN K.;
 RT "The B-cell surface protein CD72/Lyb-2 is the ligand for CDS.";
 RL Nature 351:662-665(1991).
 CC -1- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL
 CC -1- Proliferation. CDS INTERACTS WITH CD72/Lyb-2.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 SRCR DOMAINS.
 CC -1- DATABASE: NAME-PROV; NOTE-CD guide CDS entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/proc/cd/cds.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04391; CAA27979.1; -
 DR PIR: A26396; A26396.
 DR MIM; 153340; -
 DR PFAM: PF00530; SRCR; 3.
 KW Signal; Transmembrane; Glycoprotein; T-cell.
 FT CHAIN 1 24
 FT DOMAIN 25 495
 FT TRANSMEM 25 372
 FT DOMAIN 373 402
 FT DOMAIN 403 495
 FT DOMAIN 34 133
 FT DOMAIN 275 382
 FT DISULFID 44 107
 FT CARBOHYD 116 116
 CARBOHYD 241 241
 SEQUENCE 495 AA; 54625 MW; 8507C34C CRC32;
 Query Match 14.88; Score 112; DB 1; Length 495;
 Best Local Similarity 40.58; Pred. No. 5.18e-06;
 Matches 17; Conservative 11; Mismatches 12; Indels 2; Gaps 1;
 Db 277 RLVGSSICETVEVROGAQMALCDSRSSRLRWEVCRE 318
 QY 228 RLAGSSVHGRVELYHAQGMGTVCDDQMDADA--EVICRQ 267
 Search completed: Mon Mar 13 10:27:25 2000
 Job time: 10 secs.

227-3327/SPCR2
-147-947-6-04. rspt
Asc Inf Mos3

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search protein - protein database search, using Smith-Waterman algorithm
 on: Mon Mar 13 10:27:42 2000; MasPar time 16.64 Seconds
 420.878 Million cell updates/sec
 Tabular output not generated.

```

Title: >US-09-147-947-6
Description: (227-327) from US09147947A.pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLGGSSVHEGRPELYHAG.....KSSMGHHNGKHEKDAQVCT 101

```

Scoring table: PAM 150
Gap 11

```
Searched: 225878 seqs, 69334122 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

Database: `sptrembl12`

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 37.892; Variance 61.971; scale 0.611

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Alt.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	450	59.4	2043	5	096943	1	SRCR DOMIN. MEMBRANE	4.97e-83
2	427	56.4	2153	5	097375	2	SCAVENGER RECEPTOR CYS	2.00e-76
3	416	55.0	1116	4	007898	3	M30 ANTIGEN PRECURSOR	9.37e-74
4	416	55.0	1149	4	007901	4	M30 ANTIGEN, EXTRACT	9.37e-74
5	416	55.0	1151	4	007899	5	M30 ANTIGEN, CYTODIAS	9.37e-74
6	416	55.0	1156	4	007500	6	M30 ANTIGEN, CYTODIAS	9.37e-74
7	412	54.4	585	4	008380	7	MHC-2 BINDING PROTEIN	8.74e-73
8	402	53.1	578	11	P70117	8	PANCREAS CANCER-ASSOCI	2.30e-70
9	402	53.1	1785	4	099211	9	DMBT1 PROTEIN	2.30e-70
10	402	53.1	1785	4	0994V9	10	DMBT1/6RS.1 PROTEIN PR	2.30e-70
11	401	53.0	504	6	028908	11	MOCTIN (FRAGMENT)	4.01e-70
12	397	52.4	918	13	092098	12	PENNA-SRCR PROTEIN PREC	3.71e-65
13	397	52.4	1594	6	095218	13	HENSLIN.	3.71e-65
14	395	51.9	574	11	070513	14	UNKNOWN MRNA, PARTIAL	1.13e-66
15	393	52.2	356	6	076852	15	CRP-DUCITIN PRECURSOR (3.42e-66
16	385	50.9	2083	11	060997	16	PEPTIDYLPROXYL ISOMERA	2.88e-66
17	384	50.7	577	11	007797	17	SCAVENGER-RECEPTOR PRO	5.04e-66
18	383	50.6	369	6	029110	18	LYSL. OXIDASE RELATED	8.76e-66
19	383	50.6	895	13	0976M1	19	SCAVENGER RECEPTOR CYS	8.76e-66
20	383	50.6	1036	5	097376	20		8.76e-66

21	381	50.3	600	6	Q28910	MCIN (FRAGMENT).	2.65e-65
22	376	48.7	804	6	Q28113	SCAVENGER-RECEPTOR PRO	4.22e-64
23	374	48.4	480	6	Q29112	SCAVENGER-RECEPTOR PRO	1.27e-63
24	373	49.3	437	11	Q35649	CYCLOPHILIN C-ASSOCIAT	2.21e-63
25	370	48.9	576	6	Q29109	SCAVENGER-RECEPTOR PRO	1.16e-62
26	369	48.6	127	6	Q95316	MS30 ANTIGEN (FRAGMENT	2.01e-62
27	368	48.5	774	4	Q9Y4K0	LYSL OXIDASE-RELATED	3.49e-62
28	365	48.2	1290	11	Q62827	EBERLIN.	1.83e-61
29	363	48.0	546	6	Q29111	SCAVENGER-RECEPTOR PRO	5.49e-61
30	356	47.0	347	4	Q38166	SP ALPHA.	2.58e-59
31	355	46.9	518	11	Q60754	BACTERIA BINDING MACRO	4.47e-59
32	346	45.7	528	5	Q97379	SCAVENGER RECEPTOR	6.23e-57
33	346	45.7	531	5	Q17064	SEB5.	6.23e-57
34	343	45.3	638	4	Q9Y578	LYSL OXIDASE-LIKE PRO	5.22e-56
35	341	45.0	483	11	Q9W0B9	MACROPHAGE RECEPTOR MA	9.62e-56
36	339	44.8	462	6	Q28881	MEMBRANE PROTEIN SCAVE	2.97e-55
37	338	44.6	520	4	Q95853	MACROPHAGE RECEPTOR	4.98e-55
38	336	44.4	754	11	Q92175	LYSL OXIDASE-RELATED	1.48e-54
39	329	43.5	352	11	Q35301	SP-ALPHA.	6.71e-53
40	329	43.5	352	11	Q35300	SP-ALPHA.	6.71e-53
41	325	42.9	592	4	Q9Y4K7	CP6C.	5.91e-52
42	325	42.9	595	4	Q9Y4K9	CP6C.	5.91e-52
43	325	42.9	605	4	Q9Y4K8	CP6D.	5.91e-52
44	325	42.9	631	4	Q9Y4L0	CP6B.	5.91e-52
45	318	42.0	822	13	Q9YHC1	LYSL OXIDASE HOMOLOG	2.64e-50

● ALIGNMENTS

ID	RESULT
1D	PRELIMINARY; PRT: 2043 AA.
AC	096943
AC	096943:
DT	01-MAY-1999 (TREMblrel. 10, Created)
DT	01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT	01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE	SRCR DOMAIN, MEMBRANE FORM 2.
GN	SRCRM2.
OS	Gedidia cydonium (Sponge).
OC	Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC	Astrophorida; Geodidae; Geodia.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE; 98369060.
RA	BUEMBACH B., PANICER Z., DIEHL-SEIFERT B., STEFFEN R., MUENKNER J.,
RA	MUELLER I., MUELLER W.E.G.;
RT	"The putative sponge aggregation receptor. Isolation and
RT	characterization of a molecule composed of scavenger receptor
RT	cysteine-rich domains and short consensus repeats."
RL	J. Cell Sci. 111:2635-2644(1998).
DR	EMBL; Y14953; CAA75175.1; -.
DR	HSSP; P10998; IVVC.
DR	PROSITE; PS00420; SPERACT_RECEPTOR; 8.
SO	SEQUENCE 2043 AA; 220896 MW; 4CE19401 CRC32;
Query Match 59.4%; SCORE 450; DB 5; Length 2043;	
Best Local Similarity 37.0%; Pred. No. 4.9/e-82;	
Matches 57; Conservative 20; Mismatches 23; Indels 0; Gaps 0.	
Dd	IRLVGSGSPHGREGREIYYQGVGTCCDSMGOPADYVCRQLGYANASRATVAEFGRT 995 : : : : : : : : :
Oy	IRLWGSSVHGGREIELTHAGOMGVTCDDQMDADAETICQLGLSTARKMHAQYFEGS 266 ::: : : : : : : :
Dd	GEIMLDNVACITGFENSIDECRSNGWGHNCGRHDGAVC 1035 : : : : : : : :
Oy	GPVMLDEVCTGNELSTEQCPRSSWSGHHNGHKEDAGVSC 326 : : : : : : : :
RESULT	2
ID	PRELIMINARY; PRT: 2153 AA.
AC	097375
AC	097375:
DT	01-MAY-1999 (TREMblrel. 10, Created)
DT	01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT	01-NOV-1999 (TREMblrel. 12, Last annotation update)

```

DE SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.
GN SCRI2.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RA PANCER Z., RAST J., DAVIDSON E.H.;
RT "Evolution of the Immune System: Transcription Factors and Downstream
RT Genes of the Mammalian Immune System Expressed in Sea Urchin
RT Coelomocytes.";
RL Immunogenetics 0:0-0(1999).
DR EMBL: AF064259; AAD08654.1; -.
DR PROSITE: PS00420; SPERACT_RECEPTOR. 15.
KW Signal; Receptor.
FT SIGNAL 1 16
CHAIN 17 2153
SEQUENCE 2153 AA; 226557 MW; CA3BA8A9 CRC32;

Query Match 56.4%; Score 427; DB 5; Length 2153;
Best Local Similarity 55.4%; Pred. No. 2,00e-76;
Matches 56; Conservative 23; Mismatches 22; Indels 0; Gaps 0;

DB 462 VRLVGLNREGRVEIFLNNQWGTCCDDWGTDPDANVCHOLGYPGSGARSSAYFGGGS 521
OY 227 IRLAGSSVHEGRVELYHAGOMGTCCDDWDADAIEVICROLGSLAKAMHQAIFGEGS 286
DB 522 VILLIDNVCCSNGERSLSLCSNNGIGVHNGRHODASVYCT 562
OY 287 GPVWLDEVKCTGNELSTIEQCPKSSWGEHNGHREDAGVSC 327

RESULT 3
ID 007898 PRELIMINARY: PRT: 1116 AA.
AC 007898.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE M130 ANTIGEN PRECURSOR.
GN CD163.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA IAN S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z22968; CA80541.1; -.
DR EMBL: Y18388; CAB45233.1; -.
DR EMBL: Y18389; CAB45233.1; JOINED.
DR EMBL: Y18390; CAB45233.1; JOINED.
DR EMBL: Y18391; CAB45233.1; JOINED.
DR EMBL: Y18392; CAB45233.1; JOINED.
DR EMBL: Y18393; CAB45233.1; JOINED.
DR EMBL: Y18394; CAB45233.1; JOINED.
DR EMBL: Y18395; CAB45233.1; JOINED.
DR EMBL: Y18396; CAB45233.1; JOINED.
DR EMBL: Y18397; CAB45233.1; JOINED.
DR EMBL: Y18398; CAB45233.1; JOINED.
DR EMBL: Y18399; CAB45233.1; JOINED.
DR EMBL: Y18400; CAB45233.1; JOINED.
DR EMBL: Y18401; CAB45233.1; JOINED.

```

```

DR EMBL: Y18402; CAB45233.1; JOINED.
DR EMBL: Y18403; CAB45233.1; JOINED.
DR PFAM: PF00530; SRCR. 9.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
CHAIN 41 1116
SEQUENCE 1116 AA; 120979 MW; F193FBBA CRC32;

Query Match 55.0%; Score 416; DB 4; Length 1116;
Best Local Similarity 55.4%; Pred. No. 9.37e-74;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

DB 924 IRLQEGFTSCSGRVEIWHGSGWGTCCDDSDLDADAVVCOQJGCPALAKFAKFAEFGGT 983
OY 227 IRLAGSSVHEGRVELYHAGOMGTCCDDWDADAIEVICROLGSLAKAMHQAIFGEGS 286
DB 984 GPVWLDEVKCTGNELSTIEQCPKSSWGEHNGHREDAGVSC 327
OY 287 GPVWLDEVKCTGNELSTIEQCPKSSWGEHNGHREDAGVSC 327

RESULT 4
ID 007901 PRELIMINARY: PRT: 1149 AA.
AC 007901.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE M130 ANTIGEN, EXTRACELLULAR VARIANT PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA IAN S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: Z22971; CA80544.1; -.
DR PFAM: PF00530; SRCR. 9.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
CHAIN 41 1149
SEQUENCE 1149 AA; 124328 MW; 4901C708 CRC32;

Query Match 55.0%; Score 416; DB 4; Length 1149;
Best Local Similarity 55.4%; Pred. No. 9.37e-74;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

DB 957 IRLQEGFTSCSGRVEIWHGSGWGTCCDDSDLDADAVVCOQJGCPALAKFAKFAEFGGT 1016
OY 227 IRLAGSSVHEGRVELYHAGOMGTCCDDWDADAIEVICROLGSLAKAMHQAIFGEGS 286
DB 1017 GPVWLDEVKCTGNELSTIEQCPKSSWGEHNGHREDAGVSC 327
OY 287 GPVWLDEVKCTGNELSTIEQCPKSSWGEHNGHREDAGVSC 327

RESULT 5
ID 007899 PRELIMINARY: PRT: 1151 AA.
AC 007899.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.

```

[illegible]

05	Homo sapiens (Human).	
0C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Primates; Catarrhini; Homiinae; Homo.	
RN	(1)	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE: 93300818.	
RA	KOHNS K., TAYLOR E., HALENBECK R., CASPIT C., WANG A.;	
RT	"Cloning and characterization of a human Mac-2-binding protein, a new	
RT	member of the superfamily defined by the macrophage scavenger receptor	
RT	cysteine-rich domain."	
RL	J. Biol. Chem. 268:14245-14249(1993).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE: 94308070.	
RA	ULIRICH A., SUREIS I., D'EGIDO M., JALLAT B., POWELL T.J., HERBST R.;	
RA	DREPS A., AZAM M., RUBINSTEIN M., NATOI C.;	
RT	"The secreted tumor-associated antigen 90K is a potent immune	
RT	stimulator."	
RL	J. Biol. Chem. 269:18401-18407(1994).	
DR	EMBL: L13210; AAA36193.1; -	
DR	EMBL: X79089; CAA55699.1; -	
DR	PFAM: PF00530; SRCR: 1.	
DR	PRINTS: PR00258; SPBRACRCPTR.	
KW	Signal.	
FT	SIGNAL. 1 18 POTENTIAL.	
FT	CHAIN 19 585 MAC-2 BINDING PROTEIN.	
SO	SEQUENCE 585 AA; 65330 MW; AAEAE32 CRC32;	
Query Match	54.4%; Score 412; DB 4; Length 585;	
Best Local Similarity	53.5%; Pred. No. 8.74e-73;	
Matches	54; Conservative 19; Mismatches 28; Indels 0; Gaps 0	
Db	24 MRLADGATNGRGVEIEIFRGQMGVCDNLMLTASVCRALGCEFNATOLGKRAFGGS 83	
QY	227 IRLAGGSSVHGRLVLYAGQMGVCDQWDADAIVYICRLGLSGTAKMHWQYFEBS 286	
Db	84 GPVMLDEVCTGTERSLADCKSLGLMKSNCRHERDAGVCT 124	
QY	287 GPVMLDEVCTGNELSIQCPSKSWGHNCGHKEDAGVCT 327	
RESULT	8	
ID	P70117	PRELIMINARY; PRT; 578 AA.
AC	P70117	
DT	01-FEB-1997 (TREMBLrel. 02, Created)	
DT	01-FEB-1997 (TREMBLrel. 02, last sequence update)	
DT	01-NOV-1999 (TREMBLrel. 12, last annotation update)	
DE	PACCREAS CANCER-ASSOCIATED PROTEIN 4.	
OS	Mesocricetus auratus (Golden hamster).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	SCHAFFERT C., POUR P.M., MACDONALD R.G., CHANEY W.G.;	
RL	Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: U73375; AAB18745.1; -	
DR	PROSITE: PS00420; SPBRACRCPTR.	
DR	PFAM: PF00530; SRCR: 1	
DR	PRINTS: PR00258; SPBRACRCPTR.	
SO	SEQUENCE 578 AA; 64398 MW; 002EBBD2 CRC32;	
Query Match	53.1%; Score 402; DB 11; Length 578;	
Best Local Similarity	52.5%; Pred. No. 2.30e-70;	
Matches	53; Conservative 15; Mismatches 33; Indels 0; Gaps 0.	
Db	24 MRLVNGASANGRAVEIEIFRGQMGVCDNLMLTASVCRALGCEFNATOLGKRAFGGR 83	
QY	227 IRLAGGSSVHGRLVLYAGQMGVCDQWDADAIVYICRLGLSGTAKMHWQYFEBS 286	
Db	84 GPVMLDEVCTGTERSLANGSLGLMKSRCRHERDAGVCS 124	
QY	287 GPVMLDEVCTGNELSIQCPSKSWGHNCGHKEDAGVCT 327	

ID	9		PRT:	1785 AA.
AC	O9y211	PRELIMINARY;		
DT	01-NOV-1999	(TrEMBLrel. 12, Created)		
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)		
DE	01-NOV-1999	(TrEMBLrel. 12, Last annotation update)		
GN	DMBT1 PROTEIN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
CC	[1] Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
RN				
RP	HOBIIT A:			
RA	SEQUENCE FROM N.A.			
RT	"DMBT1.."			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AB020851; BAA78577.1; -			
DR	EMBL; AB020812; BAA78577.1; JOINED.			
DR	EMBL; AB020813; BAA78577.1; JOINED.			
DR	EMBL; AB020814; BAA78577.1; JOINED.			
DR	EMBL; AB020815; BAA78577.1; JOINED.			
DR	EMBL; AB020816; BAA78577.1; JOINED.			
DR	EMBL; AB020817; BAA78577.1; JOINED.			
DR	EMBL; AB020818; BAA78577.1; JOINED.			
DR	EMBL; AB020819; BAA78577.1; JOINED.			
DR	EMBL; AB020820; BAA78577.1; JOINED.			
DR	EMBL; AB020821; BAA78577.1; JOINED.			
DR	EMBL; AB020822; BAA78577.1; JOINED.			
DR	EMBL; AB020823; BAA78577.1; JOINED.			
DR	EMBL; AB020824; BAA78577.1; JOINED.			
DR	EMBL; AB020825; BAA78577.1; JOINED.			
DR	EMBL; AB020826; BAA78577.1; JOINED.			
DR	EMBL; AB020827; BAA78577.1; JOINED.			
DR	EMBL; AB020828; BAA78577.1; JOINED.			
DR	EMBL; AB020829; BAA78577.1; JOINED.			
DR	EMBL; AB020830; BAA78577.1; JOINED.			
DR	EMBL; AB020831; BAA78577.1; JOINED.			
DR	EMBL; AB020832; BAA78577.1; JOINED.			
DR	EMBL; AB020833; BAA78577.1; JOINED.			
DR	EMBL; AB020834; BAA78577.1; JOINED.			
DR	EMBL; AB020835; BAA78577.1; JOINED.			
DR	EMBL; AB020836; BAA78577.1; JOINED.			
DR	EMBL; AB020837; BAA78577.1; JOINED.			
DR	EMBL; AB020838; BAA78577.1; JOINED.			
DR	EMBL; AB020839; BAA78577.1; JOINED.			
DR	EMBL; AB020840; BAA78577.1; JOINED.			
DR	EMBL; AB020841; BAA78577.1; JOINED.			
DR	EMBL; AB020842; BAA78577.1; JOINED.			
DR	EMBL; AB020843; BAA78577.1; JOINED.			
DR	EMBL; AB020844; BAA78577.1; JOINED.			
DR	EMBL; AB020845; BAA78577.1; JOINED.			
DR	EMBL; AB020846; BAA78577.1; JOINED.			
DR	EMBL; AB020847; BAA78577.1; JOINED.			
DR	EMBL; AB020848; BAA78577.1; JOINED.			
DR	EMBL; AB020849; BAA78577.1; JOINED.			
DR	EMBL; AB020850; BAA78577.1; JOINED.			
SQ	SEQUENCE 1785 AA; 193991 MW; 479F75D8 CRC32:			

Query Match 53.1%; Score 402; DB 4; Length 1785;
 Best Local Similarity 51.5%; Pred. No. 2,30e+70;
 Matches 52; Conservative 19; Mismatches 30; Indels 0; Gaps

Dn	1258 VNLSSYGLCAGREIYHGSTGWTGVDDSDSTIOEAIVYCRQLGGCAVSALGNATFGSGS 1317
Gy	227 IRLGSSSVHGREGRIELTHAQKMTVCDDQDDADAEVIRQLGLSLGIAMWHAQAYFGGS 286
Dd	1318 GPILDDVECSGTSTLMQCRRNGGWESHNCNRREDAGVICS 1358
Gy	287 GPVMLDEVRCIGNELSTEGCPKSKSMWEHNCGHKREDAGVICT 327

RESULT 10

```

ID AC 00944V9 PRELIMINARY: PRT, 1785 AA.
DB 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE DBMT1/6KB.1 PROTEIN PRECURSOR.
GN DBMT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG.
RX MEDLINE: 97434209.
RA MOLLERHAUER J., WIEMANN S., SCHEUREN M., KORN B., HAYASHI Y.,
RA WILGENBUS K.K., VON DEMLING A., POUSTKA A.;
RT "DBMT1, a new member of the SCCR superfamily, on chromosome 10q25.3-
RT 26.1 is deleted in malignant brain tumours.";
RL Nat. Genet. 17:32-39(1997).
DR EMBL: A000342; CAA04019.1; -.
RW Signal.
FT SIGNAL. 1 25 POTENTIAL.
FT CHAIN 26 1785 DBMT1/6KB.1 PROTEIN.
SQ SEQUENCE 1785 AA; 193941 MW; 9259828E CRC32;

Query Match 53.1%; Score 402; DB 4; Length 1785;
Best Local Similarity 51.5%; Pred. No. 2.30e-70;
Matches 52; Conservative 19; Mismatches 30; Indels 0; Gaps 0

Db 1258 VNUNSSYGLCAGVEYEHGCTMGTCDDSWIQAFAVYCRQLCGRAVSAALNAYFGSGS 1317
Oy 227 IRLAGSSVHGEVGHVYHAGQGVTCDDQWDADAEVICRQLSLGIRAKMHQAYFGSGS 286
Db 1318 GPITLDVDECSGTSTLMQCRNRGWFNSHNCNRREDAGVICS 1358
Oy 287 GPVMDDEVRCGTGNELSIEGCPRSSNGEHCNRKEDAGVST 327

RESULT 11
ID 028908 PRELIMINARY: PRT, 504 AA.
AC 028908;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)
DE MUCIN (FRAGMENT).
DE Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 95374471.
RX NUNES D.P., KEATES A.C., AFDAHL N.H., OFFNER G.D.;
RT "Bovine gall-bladder mucin contains two distinct tandem repeating
RT sequences: evidence for scavenger receptor cysteine-rich repeats.";
RL Biochem. J. 310:41-48(1995).
DR EMBL: S78869; AAB35070.1; -.
DR PRAM: PF00431; CUB: 1.
DR PRAM: PF00530; SCCR: 1.
FT NON_TER 1 504
FT NON_TER 1 504
SQ SEQUENCE 504 AA; 54864 MW; EF298513 CRC32;

Query Match 53.0%; Score 401; DB 6; Length 504;
Best Local Similarity 60.4%; Pred. No. 4.01e-70;
Matches 55; Conservative 13; Mismatches 21; Indels 2; Gaps 2.

Db 328 GREVEIHGQOMCTVDDMDWDQDAQVVCRLG-CGYAVSAPGNAYFGSGSPITLDVVC 386
Oy 238 GREVEIHAQOMGTVDDQDDMDADAQVVCRLGLSLAKAMH-QAIFGSGSPVMDVVC 296
Db 387 SGAESNLMQCRNRGWFNSHNCNRREDAGVICS 417

```

[illegible]

```

Yy 238 GRVELYHAGGNGYTCDDDDMDADADAEVICRDLGSLGAKAMHQAFYFGSGSPYMLDEVICT 297
Db 1141 GTEATLMQCRSQSOWFSHNCGHHEDAVICT 1170
Yy 298 GNELSIEQCPSKSSMGNEHNCGHKEDAGVSCIT 327

RESULT 14
ID 070513 PRELIMINARY; PRT; 574 AA.
AC 070513;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE MAA.
OC Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER;
RA SHIMIZU T.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF065438; AAC17177.1;
DR PROSITE, PS00420; SPERACT_RECEPTOR; 1.
DR PFAM; PF00530; SRCR; 1.
DR PRINTS; PR00258; SPERACTRCPTR.
SQ SEQUENCE 574 AA; 63772 MW; 0EF757E3 CRC32;

Query Match 52.2%; Score 395; DB 11; Length 574;
Best Local Similarity 51.5%; Pred. No. 1,13e-68;
Matches 52; Conservative 16; Mismatches 33; Indels 0; Gaps 0.

Db 24 MRLVNGASASGREYIEFGRGWCVCNMLMDAHVCRALGENTATLSRAAFGPGK 83
Yy 227 IRLAGSSVHGRELVLHAGONGTCDDDDMDADAEVICRDLGSLGAKAMHQAFYFGS 266
Db 84 GPIMLDEVCTGNESSLANCSSLGMVNSHCGHEKDAVCS 124
Yy 287 GPVMLDEVCTGNETLSIEQCHSSMGNEHNCGHKEDAGVSCIT 327

RESULT 15
ID 097682 PRELIMINARY; PRT; 356 AA.
AC 097682;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE UNKNOWN MRNA, PARTIAL CDS (FRAGMENT).
OC Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Macaca.
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ENDOMETRIUM;
RC MEDLINE; 98439602.
RA ACE C.I., OKULICZ W.C.;
RT "A progesterone-induced endometrial homolog of a new candidate tumor
suppressor, DMB1."
RL J. Clin. Endocrinol. Metab. 83:3569-3573(1998).
DR EMBL; AF105035; AAC97514.1;
DR HSSP; P29392; ISFP.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 356 AA; 39365 MW; 949922250 CRC32;

Query Match 51.9%; Score 393; DB 6; Length 356;
Best Local Similarity 50.5%; Pred. No. 3.42e-68;
Matches 51; Conservative 22; Mismatches 28; Indels 0; Gaps 0.

Db 140 VNLSSTGLCAGREYIYHGTTGACVDSVTIEAEVVCROLGCGRAVSLGNAYFGAGS 199

```

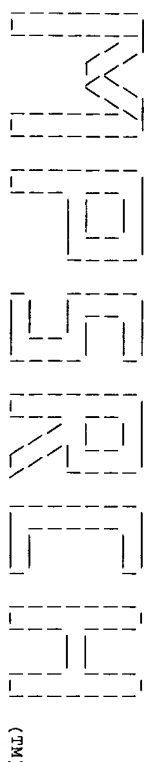
Thu Mar 16 07:57:57 2000

US-09-147-947-6-04.rsp

Page 6

Qy 227 IRLAGSSVHGRVELLYIAGOGMVICDDQMDADAELYICRDLGSLAKAMHQAYFEGS 266
Dd 200 GPITLDVECSGSESTLMQCNRNRMFSHNCRHNDAGVICS 240
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 287 GPMVLDEVRCTGNELISIQCKPSSWGEINCHKEDAGVSCT 327

Search completed: Mon Mar 13 10:28:00 2000
Job time : 18 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:25:04 2000; MasPar time 10.24 Seconds
Tabular output not generated. 465.385 Million cell updates/sec

Title: >US-09-147-947-6
(117-217) from US09147947A.pep (3 of 6)
Description: Perfect Score: 1 VRLRGKNEFEVTVYASG.....KDIWGGVCPQKMAAVTCS 101
Sequence:

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 37.928; Variance 60.892; scale 0.623

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length DB	ID	Description	Pred. No.
1	34	42.4	761	JC5759 brain-specific serine	6.42e-53
2	26	34.9	2153	JC5759 scavenger receptor cy	3.72e-39
3	25	33.5	454	I46862 macrophage scavenger	7.99e-37
4	25	32.9	1116	S36077 M130 antigen - human	1.15e-35
5	25	32.9	1149	S36077 M130 antigen (extrac	1.15e-35
6	25	32.9	1151	I38004 M130 antigen (cytosol	1.15e-35
7	25	32.9	1156	I38005 M130 antigen (cytosol	1.15e-35
8	24	32.5	454	A44407 macrophage scavenger	5.71e-35
9	24	31.9	458	B44407 macrophage scavenger	4.79e-34
10	23	31.3	518	A55840 macrophage bacteria-b	6.77e-33
11	23	31.0	451	A38415 macrophage scavenger	1.95e-32
12	26	30.9	1436	A46496 antigen WC1.1 precurs	3.31e-32
13	23	30.6	600	S56744 mucin (clone pGM7-1)	9.49e-32
14	23	30.5	453	S08276 LDL receptor I, macro	1.61e-31
15	23	30.1	918	JC4361 scavenger receptor cy	7.79e-31
16	23	28.7	585	A47161 Mac-2-binding glycopr	4.25e-28
17	21	27.7	1290	A57190 ebinerin precursor - r	9.26e-27
18	20	27.0	504	S56745 mucin (clone pGM31-1)	2.04e-25
19	20	26.8	626	S56745 muscd precursor - mu	3.41e-25
20	22	26.4	577	A53302 cyclophilin C-associat	1.59e-24
21	19	25.3	532	A32751 speract receptor prec	1.56e-22
22	18	24.6	468	S26741 T-cell glycoprotein C	1.95e-21
23	13	14.8	1035	A43090 enteropeptidase (EC 3	4.45e-06

24	110	14.4	1488	C70984	probable ppse protein	1.59e-05
25	107	14.0	1034	A53663	enteropeptidase (EC 3	5.36e-05
26	106	13.9	1489	S73015	polyketide synthase p	8.45e-05
27	105	13.7	1019	A56318	enteropeptidase (EC 3	1.28e-04
28	102	13.4	583	A29154	complement factor I (4.36e-04
29	98	12.8	495	S11270	T-cell surface glycop	2.17e-03
30	95	12.4	558	T15448	hypothetical protein	7.09e-03
31	91	11.9	442	I47074	gene CD5 protein - sh	3.32e-02
32	91	11.9	494	A29079	lymphocyte surface gl	3.32e-02
33	90	11.8	613	S15468	complement C3b/C4b in	4.86e-02
34	88	11.5	1751	MMHMH	lamina alpha-2 chain	1.03e-01
35	87	11.4	491	S58202	HMS protein - yeast	1.49e-01
36	86	11.3	527	B70920	hypothetical protein	2.16e-01
37	86	11.3	604	T15514	hypothetical protein	2.16e-01
38	86	11.3	699	T05225	extensin homolog Flr1	2.16e-01
39	84	11.0	381	T06293	3-hydroxyisobutyryl-c	4.48e-01
40	84	11.0	495	A26396	T-cell surface glycop	4.48e-01
41	83	10.9	558	UC2004	nuclear factor-kappa	6.42e-01
42	82	10.7	1113	JB0315	low-density lipoprote	9.17e-01
43	82	10.7	1719	I38902	retinoblastoma-associ	9.17e-01
44	81	10.6	1014	T13476	hypothetical protein	1.31e+00
45	81	10.6	1857	S01787	fatty-acid synthase (1.31e+00

ALIGNMENTS

RESULT	ENTRY	1
JC5759	#type complete	
JC5759	brain-specific serine protease (EC 3.4.21.-) - mouse	
JC5759	#formal_name Mus musculus #common_name house mouse	
JC5759	24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change	
JC5759	17-Mar-1999	
JC5759	Yamamura, Y.; Yamashiro, K.; Tsuruoka, N.; Nakazato, H.;	
JC5759	Tsujiimura, A.; Yamaguchi, N.	
JC5759	Biochem. Biophys. Res. Commun. (1997) 239:386-392	
JC5759	Molecular cloning of a novel brain-specific serine protease	
JC5759	with a kring-like structure and three scavenger receptor	
JC5759	cysteine-rich motifs.	
JC5759	#cross-references PMID:9808848	
JC5759	#accession	
JC5759	#molecule-type mRNA	
JC5759	#residues 1-761 #label YAM	
JC5759	#cross-references DBID:D89871	
JC5759	#experimental_source brain	
JC5759	#classification #superfamily trypsin homology; scavenger receptor	
JC5759	#feature	
JC5759	glycoprotein; hydrolase; serine proteinase	
JC5759	KEYWORDS	
JC5759	185-157	
JC5759	163-266	
JC5759	166-266, 273-372,	
JC5759	366-486	
JC5759	513-516	
JC5759	517-755	
JC5759	93,521,569	
JC5759	562,612,711	
JC5759	SUMMARY	
JC5759	length 761 #molecular-weight 84136 #checksum 5449	
JC5759	Query Match	
JC5759	Best local similarity 45.5%; Pred. No. 6.42e-53;	
JC5759	Matches 45; Conservative 19; Mismatches 34; Indels 2; Gaps 2;	
JC5759	DB 166 IRLVGNSEHGVEVTVYASG*GTCICDDQNDADADVICRGLGSIKAWHOAHF-GEQ 224	
JC5759	QY 117 VRLRGKNEFEVTVYASG*GTCICDDQNDADADVICRGLGSIKAWHOAHF-GEQ 224	
JC5759	DB 225 SGPILLDEVTRCTGNTMSIEQCRKSSWGEHNCGRKEDAGVSC 265	
JC5759	QY 117 SGPILLDEVTRCTGNTMSIEQCRKSSWGEHNCGRKEDAGVSC 265	

```
QY 176 LPIYWSNVRGRDDENLILCEKIDWGVCPQKMAAATVC 216

RESULT 2
ENTRY 114893 #type complete
TITLE scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name purple urchin
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSIONS 114893
REFERENCE 218253
AUTHORS Pancer, Z.; Rast, J.P.; Davidson, E.H.
JOURNAL Immunogenetics (1999) 49:773-786
#title Origins of immunity: transcription factors and homologs of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes.
#accession 114893
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2153 #label PAN
#cross-references EMBL:AF064259; NID:g4165052; RID:g4165053; PID:AAD0654.1

GENETICS
#gene SRCR12
SUMMARY #length 2153 #molecular-weight 226556 #checksum 3510
Query Match 34.8%; Score 266; DB 2; Length 2153;
Best Local Similarity 38.6%; Pred. No. 3,72e-39;
Matches 39; Conservative 22; Mismatches 40; Indels 0; Gaps 0;

Db 462 VRLVGLNREGRAVEIFLNNQGVCDMDGTPDANVVCROLGYPGSGARSAYFRGS 521
117 VRLRGKNEFEETVEYVAGVGTVCSSHWDDSDASVICHQLOLGKGIKQPFPSGLG 176
QY 522 VPILDNVCSSGERSLELCSNNGIGVHNGCHQEDASVYCT 562
177 IPIYWSNVRGRDDENLILCEKIDWGVCPQKMAAATVC 217

RESULT 3
ENTRY 146862 #type complete
TITLE macrophage scavenger receptor type I - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 20-Sep-1999
ACCESSIONS 146862
REFERENCE 146862
AUTHORS Bickel, P.E.; Freeman, M.W.
JOURNAL J. Clin. Invest. (1992) 90:1450-1457
#title Rabbit aortic smooth muscle cells express inducible macrophage scavenger receptor messenger RNA that is absent from endothelial cells.
#cross-references MUID:93016877
#accession 146862
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-454 #label BIC
#cross-references GB:111933; NID:g165508; PID:AAA31402.1; RID:g165509
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor cysteine-rich domain homology
FEATURE 350-453 #domain scavenger receptor cysteine-rich domain homology
SUMMARY #length 454 #molecular-weight 49745 #checksum 2903
Query Match 33.5%; Score 256; DB 2; Length 454;
Best Local Similarity 40.8%; Pred. No. 7.99e-37;
Matches 42; Conservative 17; Mismatches 40; Indels 4; Gaps 4;

Db 353 VRLVGLNREGRAVEIFLNNQGVCDMDGTPDANVVCROLGYPGSGARSAYFRGS 521
177 IPIYWSNVRGRDDENLILCEKIDWGVCPQKMAAATVC 217

QY 117 VRLRGKNEFEETVEYVAGVGTVCSSHWDDSDASVICHQLOLGKGIKQPFPSGLG 175
412 TGPI-WLNEVPCLGMSIECKIRQWGVRCSHGEDAGVCT 453
QY 176 LPIYWSNVRGRDDENLILCEKIDWGVCPQKMAAATVC 217

RESULT 4
ENTRY 536077 #type complete
TITLE M30 antigen - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Aug-1996
ACCESSIONS 138003; 536077
REFERENCE 138003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
JOURNAL Eur. J. Immunol. (1993) 23:2320-2325
#title A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
#accession 138003
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1116 #label RES
#cross-references EMBL:Z22968; NID:g312141; PID:g312142
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 43-147
151-254 #domain scavenger receptor cysteine-rich domain homology
258-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology
470-573 #domain scavenger receptor cysteine-rich domain homology
575-678 #domain scavenger receptor cysteine-rich domain homology
711-814 #domain scavenger receptor cysteine-rich domain homology
816-920 #domain scavenger receptor cysteine-rich domain homology
921-1024 #domain scavenger receptor cysteine-rich domain homology
SUMMARY #length 1116 #molecular-weight 120979 #checksum 5687
Query Match 32.9%; Score 251; DB 2; Length 1116;
Best Local Similarity 39.8%; Pred. No. 1.15e-35;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;

Db 924 IRLQEGPTCSGRVETWAGSGWTCVDSWDLDPAQVYCOQLGCGPALKAKRAEF-GGG 982
117 VRLRGKNEFEETVEYVAGVGTVCSSHWDDSDASVICHQLOLGKGIKQPFPSGLG 175
QY 983 TGPI-WLNEVCKGNSSSLMDCPARWGHSEGHKEDAAVCT 1024
176 LPIYWSNVRGRDDENLILCEKIDWGVCPQKMAAATVC 217

RESULT 5
ENTRY 138006 #type complete
TITLE M130 antigen (extracellular variant) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Aug-1996
ACCESSIONS 138006; 535768
REFERENCE 138003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
JOURNAL Eur. J. Immunol. (1993) 23:2320-2325
```

```
#title      A new macrophage differentiation antigen which is a member of
#cross-references MUID:93380506
#accession  I38006
#status     Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-1149 ##label RES
#cross-references EMBL:222971; NID:g312147; PID:g312148
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE
43-147      #domain scavenger receptor cysteine-rich domain homology
151-254     #domain scavenger receptor cysteine-rich domain homology
258-361     #domain scavenger receptor cysteine-rich domain homology
468         #domain scavenger receptor cysteine-rich domain homology
508-711     #domain scavenger receptor cysteine-rich domain homology
744-847     #domain scavenger receptor cysteine-rich domain homology
849-953     #domain scavenger receptor cysteine-rich domain homology
954-1057    #domain scavenger receptor cysteine-rich domain homology
SUMMARY     #length 1149 #molecular-weight 124328 #checksum 487

Query Match      32.9%; Score 251; DB 2; Length 1149;
Best Local Similarity 39.8%; Pred. No. 1.15e-35;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;

ACCESSIONS
REFERENCE
#authors      Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
               Willis, A.C.; Mason, D.Y.
#journal      Eur. J. Immunol. (1993) 23:2320-2325
#title        A new macrophage differentiation antigen which is a member of
               the scavenger receptor superfamily.
#cross-references MUID:93380506
#accession    I38004
#status       Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-1151 ##label RES
#cross-references EMBL:222969; NID:g312143; PID:g312144
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS      cytosol
FEATURE
43-147      #domain scavenger receptor cysteine-rich domain homology
151-254     #domain scavenger receptor cysteine-rich domain homology
258-361     #domain scavenger receptor cysteine-rich domain homology
365-468     #domain scavenger receptor cysteine-rich domain homology
921-1024    #domain scavenger receptor cysteine-rich domain homology
SUMMARY     #length 1156 #molecular-weight 125352 #checksum 9043

Query Match      32.9%; Score 251; DB 2; Length 1156;
Best Local Similarity 39.8%; Pred. No. 1.15e-35;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
```

```
470-573     #domain scavenger receptor cysteine-rich domain homology
575-678     #domain scavenger receptor cysteine-rich domain homology
711-814     #domain scavenger receptor cysteine-rich domain homology
816-920     #domain scavenger receptor cysteine-rich domain homology
921-1024    #domain scavenger receptor cysteine-rich domain homology
SUMMARY     #length 1151 #molecular-weight 124820 #checksum 1481

Query Match      32.9%; Score 251; DB 2; Length 1151;
Best Local Similarity 39.8%; Pred. No. 1.15e-35;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;

ACCESSIONS
REFERENCE
#authors      Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
               Willis, A.C.; Mason, D.Y.
#journal      Eur. J. Immunol. (1993) 23:2320-2325
#title        A new macrophage differentiation antigen which is a member of
               the scavenger receptor superfamily.
#cross-references MUID:93380506
#accession    I38005
#status       Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-1156 ##label RES
#cross-references EMBL:222970; NID:g312145; PID:g312146
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS      cytosol
FEATURE
43-147      #domain scavenger receptor cysteine-rich domain homology
151-254     #domain scavenger receptor cysteine-rich domain homology
258-361     #domain scavenger receptor cysteine-rich domain homology
365-468     #domain scavenger receptor cysteine-rich domain homology
470-573     #domain scavenger receptor cysteine-rich domain homology
575-678     #domain scavenger receptor cysteine-rich domain homology
711-814     #domain scavenger receptor cysteine-rich domain homology
816-920     #domain scavenger receptor cysteine-rich domain homology
921-1024    #domain scavenger receptor cysteine-rich domain homology
SUMMARY     #length 1156 #molecular-weight 125352 #checksum 9043

Query Match      32.9%; Score 251; DB 2; Length 1156;
Best Local Similarity 39.8%; Pred. No. 1.15e-35;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;

ACCESSIONS
REFERENCE
#authors      Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
               Willis, A.C.; Mason, D.Y.
#journal      Eur. J. Immunol. (1993) 23:2320-2325
#title        A new macrophage differentiation antigen which is a member of
               the scavenger receptor superfamily.
#cross-references MUID:93380506
#accession    I38005
#status       Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-1156 ##label RES
#cross-references EMBL:222970; NID:g312145; PID:g312146
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS      cytosol
FEATURE
43-147      #domain scavenger receptor cysteine-rich domain homology
151-254     #domain scavenger receptor cysteine-rich domain homology
258-361     #domain scavenger receptor cysteine-rich domain homology
365-468     #domain scavenger receptor cysteine-rich domain homology
470-573     #domain scavenger receptor cysteine-rich domain homology
575-678     #domain scavenger receptor cysteine-rich domain homology
711-814     #domain scavenger receptor cysteine-rich domain homology
816-920     #domain scavenger receptor cysteine-rich domain homology
921-1024    #domain scavenger receptor cysteine-rich domain homology
SUMMARY     #length 1156 #molecular-weight 125352 #checksum 9043

Query Match      32.9%; Score 251; DB 2; Length 1156;
Best Local Similarity 39.8%; Pred. No. 1.15e-35;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
```

QY 117 VRLRGKNEFEGETVEYVAGWGTVCSSHMDDSDASVICHOLQLOGKGLA-KQTPFSGLG 175

Db 983 TGPI-WLNEVCKCKNGESSLMDCPARRMGHSECGHAKDAVNC 1024

QY 176 LIPIYW-SNVRGRGDEENILLCEKDIWOGVCPQKMAAVTCS 217

RESULT 8

ENTRY A44407 #type complete

TITLE macrophage scavenger receptor, RSRI (collagen-like domain) - rabbit

ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit

DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999

ACCESSIONS A44407

REFERENCE A44407

authors Dol, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.; Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.; Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.

Journal J. Biol. Chem. (1993) 268:2126-2133

Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors.

#cross-references MUID:93131972

#accession A44407

#status preliminary; not compared with conceptual translation

#molecule_type nucleic acid

#residues 1-454 #label DOI

#experimental_source lung

#note sequence extracted from NCBI backbone (NCBI:P123206)

CLASSIFICATION #superfamily unassigned collagens; scavenger receptor

FEATURE 350-453

350-453 #domain scavenger receptor cysteine-rich domain homology

SUMMARY #label SRC

#length 454 #molecular-weight 49735 #checksum 2435

Query Match 32.5%; Score 248; DB 2; Length 454;

Best Local Similarity 39.8%; Pred. No. 5,71e-35;

Matches 41; Conservative 17; Mismatches 41; Indels 4; Gaps 4;

Db 353 VRLVGGSGHGRVREILHNGMGVLCDDHWEIRAGOVCRSLGTVGSVAKKAF-GQG 411

QY 117 VRLRGKNEFEGETVEYVAGWGTVCSSHMDDSDASVICHOLQLOGKGLA-KQTPFSGLG 175

Db 412 TGPI-WLNEVPCLEMESSIECKIRMGVRCVSHGEDAGVCT 453

QY 176 LIPIYW-SNVRGRGDEENILLCEKDIWOGVCPQKMAAVTCS 217

ENTRY 9

TITLE B44407 #type complete

ALTERNATE_NAMES macrophage scavenger receptor, MSRI - mouse

ORGANISM #formal_name Mus musculus #common_name house mouse

DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999

ACCESSIONS B44407; A38260; I56334

REFERENCE A44407

authors Dol, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.; Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.; Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.

Journal J. Biol. Chem. (1993) 268:2126-2133

Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors.

#cross-references MUID:93131972

#accession B44407

#status preliminary; not compared with conceptual translation

#molecule_type nucleic acid

#residues 1-458 #label DOI

#experimental_source macrophage-like cell line P388D

REFERENCE #note

A38260 sequence extracted from NCBI backbone (NCBI:P123207)

authors Freeman, M.; Ashkenas, J.; Rees, D.J.G.; Kingsley, D.M.; Copeland, N.G.; Jenkins, N.A.; Krieger, M.

Journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8810-8814

title An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors.

#cross-references MUID:91062370

#accession A38260

#status preliminary

#molecule_type mRNA

#residues 349-458 #label FRE

#cross-references GB:M59445; GB:M36817; NID:g192737; PIDN:AAA37464.1; PID:g192738

REFERENCE I56334

authors Ashkenas, J.; Penman, M.; Vassile, E.; Acton, S.; Freeman, M.W.; Krieger, M.

Journal J. Lipid Res. (1993) 34:983-1000

title Structures and high and low affinity ligand binding properties of murine type I and type II macrophage scavenger receptors.

#cross-references MUID:93359822

#accession I56334

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type mRNA

#residues 5-458 #label RES

#cross-references GB:I04274; NID:g293745; PIDN:AAA39747.1; PID:g293746

CLASSIFICATION #superfamily unassigned collagens; scavenger receptor

FEATURE 354-457

354-457 #domain scavenger receptor cysteine-rich domain homology

SUMMARY #label SRC

#length 458 #molecular-weight 50130 #checksum 1435

Query Match 31.9%; Score 244; DB 2; Length 458;

Best Local Similarity 36.9%; Pred. No. 4,79e-34;

Matches 38; Conservative 19; Mismatches 42; Indels 4; Gaps 4;

Db 357 VRLVGGSGHGRVREILHNGMGVLCDDHWEIRAGOVCRSLGTVGSVAKKAF-GQG 415

QY 117 VRLRGKNEFEGETVEYVAGWGTVCSSHMDDSDASVICHOLQLOGKGLA-KQTPFSGLG 175

Db 416 TGPI-WLNEVPCLEMESSIECKIRMGVRCVSHGEDAGVCT 457

QY 176 LIPIYW-SNVRGRGDEENILLCEKDIWOGVCPQKMAAVTCS 217

RESULT 10

ENTRY A55840 #type complete

TITLE macrophage bacteria-binding receptor MARCO - mouse

ORGANISM #formal_name Mus musculus #common_name house mouse

DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 16-Jul-1999

ACCESSIONS A55840

REFERENCE A55840

authors Elomaa, O.; Kangas, M.; Sahlberg, C.; Tuukkanen, J.; Sormunen, R.; Liakka, A.; Thesleff, I.; Kraal, G.; Tryggvason, K.

Journal Cell (1995) 80:603-609

title Cloning of a novel bacteria-binding receptor structurally related to scavenger receptors and expressed in a subset of macrophages.

#cross-references MUID:95171455

#accession A55840

#status preliminary

#molecule_type mRNA

#residues 1-518 #label ELO

#cross-references GB:U18424; NID:g682722; PID:g682723

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology

KEYWORDS transmembrane protein

FEATURE 418-518

418-518 #domain scavenger receptor cysteine-rich domain homology

SUMMARY #label SRC
#length 518 #molecular-weight 52730 #checksum 8159

Query Match 31.3%; Score 239; DB 2; Length 518;
Best Local Similarity 39.2%; Pred. No. 6, 77e-33;

Matches 40; Conservative 21; Mismatches 34; Indels 7; Gaps 6;

DB 423 VRIMGSTNF--GRAEYNNWGTICDDMDNDATVFCRLMGLYS-RGRA-LSSYGG-GS 477

117 VRLRGKNEFEQVEYVYASGVWGTVCSSHWDDSDASVICHQQLGKGIKAKQFPFGDL 176

DB 478 GNI-WLDVNCRGTESSMDCKSNMNCVHNEDAGYECSS 518

177 IPIYW-SNVRGDEENILLCERDIWGGVCPQKMAAAVTCSS 217

11

A38415 #type complete
macrophage scavenger receptor splice form I - human

ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 20-Sep-1999

ACCESSIONS A38415, A44408

REFERENCE A38415
#authors Matsumoto, A.; Naito, M.; Itakura, H.; Ikemoto, S.; Asaoka, H.; Hayakawa, I.; Kanamori, H.; Aburatani, H.; Takaku, F.; Suzuki, H.; Kobari, Y.; Miyai, T.; Takahashi, K.; Cohen, E. H.; Wydro, R.; Housman, D. E.; Kodama, T.

#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9133-9137
#title Human macrophage scavenger receptors: primary structure, expression, and localization in atherosclerotic lesions.

#cross-references MUID:91067661

#accession A38415

##status preliminary

##molecule_type mRNA

##residues 1-451 #label MAT

##cross-references GB:D90187; NID:g219989; PIDN:BAAL4208.1;

PID:d1014913; PID:g219990

REFERENCE A44408
#authors Emi, M.; Asaoka, H.; Matsumoto, A.; Itakura, H.; Kurihara, Y.; Wada, Y.; Kanamori, H.; Itazaki, Y.; Takahashi, E.; Lepert, M.; Lalouel, J.; Kodama, T.; Mukai, T.

#journal J. Biol. Chem. (1993) 268:2120-2125
#title Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene.

#cross-references MUID:93131971

#accession A44408

##status preliminary; not compared with conceptual translation

##molecule_type nucleic acid

##residues 155-272 #label EMI

##note sequence extracted from NCBI backbone (NCBIF:123189)

GENETICS GDB:MSRI

##cross-references GDB:128046; OMIM:153622

#map_position 8p22-8p22

CLASSIFICATION #superfamily unassigned collagens; scavenger receptor

KEYWORDS #domain scavenger receptor cysteine-rich domain homology

FEATURE 347-450 #domain scavenger receptor cysteine-rich domain homology

SUMMARY #length 451 #molecular-weight 49762 #checksum 6168

Query Match 31.0%; Score 237; DB 2; Length 451;
Best Local Similarity 37.9%; Pred. No. 1, 95e-32;

Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHERVETLHSGWGTICDDREYRVGVGVCRSRGYPVQVHKAHF-GGG 408

117 VRLRGKNEFEQVEYVYASGVWGTVCSSHWDDSDASVICHQQLGKGIKAKQFPFGDL 175

DB 409 TSPPI-WLNEVFCFGRESSIECKIRQWGTIRACSHSEDACVCT 450

OY 176 IPIYW-SNVRGDEENILLCERDIWGGVCPQKMAAAVTCSS 217

RESULT 12

ENTRY A46496 #type complete

TITLE #formal_name Bos primigenius taurus #common_name cattle

ORGANISM #formal_name Bos primigenius taurus #common_name cattle

DATE 18-Jun-1993 #sequence_revision 19-May-1994 #text_change 13-Nov-1998

ACCESSIONS A46496, I45834, S19913

REFERENCE A46496
#authors Wijngaard, P.L.; Metzelaar, M.J.; Machugh, N.D.; Morrison, W.L.; Clevers, H.C.

#journal J. Immunol. (1992) 149:3273-3277
#title Molecular characterization of the WCL antigen expressed specifically on bovine CD4-CD8- gamma delta T lymphocytes.

#cross-references MUID:93056489

#accession A46496

##status preliminary

##molecule_type mRNA

##residues 1-1436 #label WJ

##cross-references EMBL:X63723; NID:g13; PID:g14

##experimental_source CD4-CD8- gamma delta T lymphocytes

##note Sequence extracted from NCBI backbone (NCBIF:117475)

REFERENCE I45834
#authors Wijngaard, P.L.; Machugh, N.D.; Metzelaar, M.J.; Romberg, S.; Bensard, A.; Pepin, L.; Davis, W.C.; Clevers, H.C.

#journal J. Immunol. (1994) 152:3476-3482
#title Members of the novel WCL gene family are differentially expressed on subsets of bovine CD4-CD8- gamma delta T lymphocytes.

#cross-references MUID:94194107

#accession I45834

##status preliminary; translated from GB/EMBL/DBJ

##molecule_type mRNA

##residues 1-1436 #label WJ2

##cross-references EMBL:X63723; NID:g13; PID:g14

GENETICS WCL.1

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology

FEATURE 1-23 #domain signal sequence #status predicted #label SIG

24-1436 #product WCL antigen #status predicted #label MAT

25-131 #domain scavenger receptor cysteine-rich domain homology

132-234 #label SR01\

236-340 #domain scavenger receptor cysteine-rich domain homology

373-476 #label SR02\

478-581 #domain scavenger receptor cysteine-rich domain homology

583-686 #label SR05\

667-789 #domain scavenger receptor cysteine-rich domain homology

791-895 #label SR06\

928-1031 #domain scavenger receptor cysteine-rich domain homology

1033-1136 #label SR09\

1152-1255 #domain scavenger receptor cysteine-rich domain homology

SUMMARY #length 1436 #molecular-weight 154196 #checksum 7163

Query Match 30.9%; Score 236; DB 2; Length 1436;
Best Local Similarity 37.9%; Pred. No. 3, 31e-32;

Matches 39; Conservative 20; Mismatches 40; Indels 4; Gaps 4;

DB 1155 LRLRGDESGRVEVWNGSVGTCDSDWSLAEVYVCOQAGCOALEVNSAAF-GPG 1213

117 VRLRGKNEFEETVEYASGVWGTVCSSHWDDSDASVICHQLOLQ-GKGIKOTPPSGLG 175
Db 1214 NGSI-WLDEYOCGRRESSLMDCVAEPMGSDCKHEPDAGVRC 1255
176 LPIYV-SNVRGRGDEENILLCCKDIMOGGVCPQKMAAVTCS 217

RESULT 13
ENTRY S56744 #type fragment
TITLE mucin (clone PGM7-1) - bovine (fragment)
ORGANISM #formal name Bos primigenius taurus #common name cattle
DATE 27-Oct-1995 #sequence-revision 30-Jan-1998 #text-change 07-May-1999

ACCESSIONS S56744
REFERENCE S56744
#authors Nunes, D.P.; Keates, A.C.; Afshal, N.H.; Offner, G.D.
#journal Biochem. J. (1995) 310:41-48
#title Bovine gall-bladder mucin contains two distinct tandem repeating sequences: evidence for scavenger receptor cysteine-rich repeats.
#cross-references MUID:95374471
#accession S56744
#molecule-type mRNA
#residues 1-600 #label NUN

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 1-102
126-229 #domain scavenger receptor cysteine-rich domain homology (fragment) #label SRC1\
253-356 #domain scavenger receptor cysteine-rich domain homology #label SRC2\
380-483 #domain scavenger receptor cysteine-rich domain homology #label SRC3\
#domain scavenger receptor cysteine-rich domain homology #label SRC4

SUMMARY #length 600 #checksum 7824

Query Match 30.6%; Score 234; DB 2; Length 600;
Best Local Similarity 38.2%; Pred. No. 9,49e-32;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

Db 383 LRLVNGSDRCGRGFEVLGGSMGTCDSDMTDNANVVCRGSGSISAPGDARF-GGG 441
QY 117 VRLRGKNEFEETVEYASGVWGTVCSSHWDDSDASVICHQLOLQ-GKGIKOTPPSGLG 175
Db 442 SGPIILDVCGSGYETLYMSCSHSPWNSHCHSKDASYICS 483
QY 176 LPIYVSNVRGRGDEENILLCCKDIMOGGVCPQKMAAVTCS 217

14
S08276 #type complete
LDL receptor I, macrophage - bovine
#formal name Bos primigenius taurus #common name cattle
DATE 30-Jun-1992 #sequence-revision 30-Jun-1992 #text-change 20-Sep-1999

ACCESSIONS S08276
REFERENCE S08276
#authors Kodama, T.; Freeman, M.; Rohrer, L.; Zabrecky, J.; Matsuda, P.; Krieger, M.
#journal Nature (1990) 343:531-535
#title Type I macrophage scavenger receptor contains alpha-helical and collagen-like coiled coils.
#cross-references MUID:90136965
#accession S08276
#status not compared with conceptual translation
#molecule-type mRNA
#residues 1-453 #label KOD
#cross-references GB:X51689; GB:X54182; NID:9734; PIDN:CA35987.1; PID:9735

CLASSIFICATION #superfamily unassigned collagens; scavenger receptor

KEYWORDS cysteine-rich domain homology
alternative splicing; coiled coil; glycoprotein; transmembrane protein
FEATURE 51-76
349-452 #domain transmembrane #status predicted #label TMN\
#domain scavenger receptor cysteine-rich domain homology #label SRC\
82,101,142,183,220, #binding-site carbohydrate (Asn) (covalent) #status predicted
248,266

SUMMARY #length 453 #molecular-weight 50056 #checksum 4303

Query Match 30.5%; Score 233; DB 2; Length 453;
Best Local Similarity 37.9%; Pred. No. 1.61e-31;
Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

Db 352 VRLVGSGGHEERVELFPHGOMGTCDRWEIRGLVCRSLGXYGVSKRAVYF-GKG 410
QY 117 VRLRGKNEFEETVEYASGVWGTVCSSHWDDSDASVICHQLOLQ-GKGIKOTPPSGLG 175
Db 411 TGPI-WLNEVFCGRKSSIEECRIQMGVRCASHDEDAVTC 452
QY 176 LPIYV-SNVRGRGDEENILLCCKDIMOGGVCPQKMAAVTCS 217

RESULT 15
ENTRY JC4361 #type complete
TITLE scavenger receptor Cys-rich epidermal growth factor precursor - sea lamprey
ORGANISM #formal name Petromyzon marinus #common name sea lamprey
DATE 08-Jan-1996 #sequence-revision 08-Feb-1996 #text-change 16-Jul-1999

ACCESSIONS JC4361
REFERENCE JC4361
#authors Mayer, W.E.; Tichy, H.
#journal Gene (1995) 164:267-271
#title A cDNA clone from the sea lamprey Petromyzon marinus coding for a scavenger receptor Cys-rich (SRCR) domain protein.
#cross-references MUID:96069593
#accession JC4361

CLASSIFICATION #superfamily EGF homology; scavenger receptor cysteine-rich domain homology
glycoprotein; growth factor; receptor; transmembrane protein
KEYWORDS
FEATURE 1-24
25-918
55-157
179-212
234-267
295-328
356-389
410-443
463-565
876-895
896-918
40,545,575,585,
814

SUMMARY #length 918 #molecular-weight 101417 #checksum 2936

Query Match 30.1%; Score 230; DB 2; Length 918;
Best Local Similarity 39.8%; Pred. No. 7.79e-31;
Matches 41; Conservative 17; Mismatches 40; Indels 5; Gaps 5;

Thứ Năm 16 07:57:49 2000

US-09-147-947-6-03.rpr

Page 7

[illegible]

Search completed: Mon Mar 13 10:25:18 2000
Job time : 14 secs.

THIS PAGE BLANK (03PT0)


```
CC      SEQUENCE CHARACTERISTICS:  
CC      LENGTH: 489 amino acids  
CC      TYPE: amino acid  
CC      STRANDEDNESS: Single  
CC      TOPOLOGY: linear  
CC      MOLECULE TYPE: protein  
CC      SEQUENCE 489 AA; 49441 MW; 1141296 CN;  
  
Query Match          31.3%; Score 239; DB 2; Length 489;  
Best Local Similarity 39.2%; Pred. No. 6,02e-15;  
Matches 40; Conservative 21; Mismatches 34; Indels 7; Gaps 6.  
  
Db      394 VRIMGTR--GRAEYVNNENGTCDDDDQMNDAIVCGRLGYS-RGRALSSYG-GS 448  
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Oy      117 VRLRGKHEFEGETVEVYASGVWGIVCSSHWDSDASVYCHQLQGKGIAIQTFFSGIGL 176  
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db      449 GNI-WLDNVNCRGTENSLMDCSKNSWGNHCNVHNEDAGVECS 489  
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
        177 IPIYW-SNVRCRDEBITLLCEKDIWQGVCPQKKAAAVTCS 217  
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
        KGGPDT  
ID      US-08-392-367B-2 STANDARD; PRT; 518 AA.  
XX      2  
AC      xxxxxx  
XT  
DT  
XX  
XX  
XX  
Sequence 2, Application US/08392367B  
DE  
CC      Sequence 2, Application US/08392367B  
CC      Patent No. 5691197  
CC      GENERAL INFORMATION:  
CC      APPLICANT: Tryggevason, Karl  
CC      APPLICANT: Elomaa, Outi  
CC      APPLICANT: Kangas, Maarit  
CC      TITLE OF INVENTION: An Insolated DNA Sequence For a  
CC      Patent No. 5691197  
CC      TITLE OF INVENTION: No. 5691197el Macrophage Receptor with  
CC      TITLE OF INVENTION: a Collagenous Domain and the  
CC      TITLE OF INVENTION: Polypeptide Chain Encoded by  
CC      TITLE OF INVENTION: such a Sequence  
CC      NUMBER OF SEQUENCES: 2  
CC      CORRESPONDENCE ADDRESS:  
CC      ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
CC      ADDRESSEE: Minnich & McKee  
CC      STREET: 1100 Superior Avenue  
CC      STREET: Suite 700  
CC      City: Cleveland  
CC      STATE: Ohio  
CC      COUNTRY: U.S.A.  
CC      ZIP: 44114-2518  
CC      COMPUTER READABLE FORM:  
CC      MEDIUM TYPE: Diskette, 3.50 inch,  
CC      MEDIUM TYPE: 720 kb storable  
CC      COMPUTER: IBM PS/2, Model 35 SX  
CC      OPERATING SYSTEM: DOS 5.0  
CC      SOFTWARE: Word Perfect 5.1  
CC      CURRENT APPLICATION DATA:  
CC      APPLICATION NUMBER: US/08/392,367B  
CC      FILING DATE:  
CC      CLASSIFICATION: 435  
CC      ATTORNEY/AGENT INFORMATION:  
CC      NAME: Minnich, Richard J.  
CC      REGISTRATION NUMBER: 24,175  
CC      REFERENCE/DOCKET NUMBER: TRV 2 009  
CC      TELECOMMUNICATION INFORMATION:  
CC      TELEPHONE: (216) 861-5582  
CC      TELEFAX: (216) 241-1666  
CC      TELEX: (216) 980162  
CC      INFORMATION FOR SEO ID NO.: 2:  
CC      SEQUENCE CHARACTERISTICS:  
CC      LENGTH: 518 amino acids
```

[illegible]

QY 176 LIPIYW-SNVRCRGDEENILLCERDIWOGVCPQKMAAAYTCS 217

RESULT 4 STANDARD: PRT: 451 AA.

AC xxxxxx

DE Sequence 2, Application US/08453117

CC Sequence 2, Application US/08453117

CC Patent No. 5683903

CC GENERAL INFORMATION:

CC APPLICANT: Lysko, Paul G.

CC APPLICANT: Elshourbagy, Nabil A.

CC TITLE OF INVENTION: Attachment Enhanced 293 Cells

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Smithkline Beecham - Corporate Patents

CC ADDRESSEE: U.S.

CC STREET: Mailcode - UW2220, 709 Swedeland Road

CC CITY: King of Prussia

CC STATE: Pennsylvania

CC COUNTRY: U.S.A.

CC ZIP: 19406-5090

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/453,117

CC FILING DATE:

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Jervais, Herbert H.

CC REGISTRATION NUMBER: 31,171

CC REFERENCE/DOCKET NUMBER: SBC-P50338

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (610) 270-5019

CC TELEFAX: (610) 270-5090

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 451 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;

Query Match 31.0%; Score 237; DB 1; Length 451;

Best Local Similarity 37.9%; Pred. No. 9,70e-15;

Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHRGREYELHSGQWGTICDDREVRVGOVCRSLGYPGVAVKAAHF-GQG 408

QY 117 VRLGKNGKEFEQTEYVYASVWGTVCSSHWDSDASVICHQLQLGG-KGIKQTPFSGLG 175

DB 409 TGPI-WLNEVFCRGRESIECKIRQWGTACSHSEDAVYCT 450

QY 176 LIPIYW-SNVRCRGDEENILLCERDIWOGVCPQKMAAAYTCS 217

RESULT 5 STANDARD: PRT: 451 AA.

AC xxxxxx

DE Sequence 2, Application US/08948222

XX Sequence 2, Application US/08948222

CC Patent No. 5683798

CC GENERAL INFORMATION:

CC APPLICANT: Lysko, Paul G.

CC APPLICANT: Elshourbagy, Nabil A.

CC TITLE OF INVENTION: Attachment Enhanced 293 Cells

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Smithkline Beecham - Corporate Patents

CC ADDRESSEE: U.S.

CC STREET: Mailcode - UW2220, 709 Swedeland Road

CC CITY: King of Prussia

CC STATE: Pennsylvania

CC COUNTRY: U.S.A.

CC ZIP: 19406-5090

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/948,222

CC FILING DATE:

CC CLASSIFICATION:

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US/08/453,117

CC FILING DATE:

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Jervais, Herbert H.

CC REGISTRATION NUMBER: 31,171

CC REFERENCE/DOCKET NUMBER: SBC-P50338

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (610) 270-5019

CC TELEFAX: (610) 270-5090

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 451 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;

Query Match 31.0%; Score 237; DB 2; Length 451;

Best Local Similarity 37.9%; Pred. No. 9,70e-15;

Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHRGREYELHSGQWGTICDDREVRVGOVCRSLGYPGVAVKAAHF-GQG 408

QY 117 VRLGKNGKEFEQTEYVYASVWGTVCSSHWDSDASVICHQLQLGG-KGIKQTPFSGLG 175

DB 409 TGPI-WLNEVFCRGRESIECKIRQWGTACSHSEDAVYCT 450

QY 176 LIPIYW-SNVRCRGDEENILLCERDIWOGVCPQKMAAAYTCS 217

RESULT 6 STANDARD: PRT: 451 AA.

AC xxxxxx

DE Sequence 2, Application US/08973145

CC Sequence 2, Application US/08973145

CC Patent No. 5919636

CC GENERAL INFORMATION:

CC APPLICANT: Lysko, Paul G.

CC APPLICANT: Elshourbagy, Nabil A.

CC APPLICANT: Brawner, Mary E.

CC TITLE OF INVENTION: Attachment Enhanced 293 Cells


```
SO SEQUENCE 491 AA; 54317 MW; 1350575 CN;  
Dd LENGTH: 453  
  
Query Match          29.3%; Score 224; DB 4; Length 453;  
Best Local Similarity 36.9%; Pred. No. 2,13e-13;  
Matches              38; Conservative 18; Mismatches 43; Indels 4
```

Dd 352 VRLVGGSGPHGRGVYFIHFGOMGATVCDDRWELRGVLVCRSLGYKQVSVKRAVA
||| || | : | : | : | : | : | : | : | : | : | :
Oy 117 VLRLGKGKEFEFVEIVASGLWVGICSSHWDSDSAVTCHOLQ-GKKGIAROTHT
|| | : | : | : | : | : | : | : | : | : | : | :
Db 411 TGPI-WLNEVFCEFPRESSIEECRIQMVRACASHDEDAGVCT 452
|| | : | : | : | : | : | : | : | : | : | : | :
Oy 176 LIPITW-SNVMCGRGEENILLCCENDIMOGCVCPQRMAAAYICS 217
||| || | : | : | : | : | : | : | : | : | : | :

T 9
US-08-316-714-10 STANDARD; PRT; 585 AA.
XXXXXX

Xx Sequence 10, Application US/08316714
Xx DE Patent No. 5965382
Xx CC GENERAL INFORMATION:
Cc APPLICANT: Kolts, Kirston E.
Cc APPLICANT: Halenbeck, Robert F.
Cc APPLICANT: Taylor, Eric W.
Cc APPLICANT: Wang, Alice M.
Cc APPLICANT: Caspipl, Clayton L.
Cc TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
Cc NUMBER OF SEQUENCES: 11
Cc CORRESPONDENCE ADDRESS:
Cc ADDRESSEE: Celus Oncology Corporation
Cc STREET: 1400 Fifty-Third Street
Cc CITY: Emeryville
Cc STATE: CA
Cc COUNTRY: USA
Cc ZIP: 94608
Cc COMPUTER READABLE FORM:
Cc MEDIUM TYPE: Floppy disk
Cc COMPUTER: IBM PC compatible
Cc OPERATING SYSTEM: PC-DOS/MS-DOS
Cc SOFTWARE: PatentIn Release #1.0, Version #1.25
Cc CURRENT APPLICATION DATA:
Cc APPLICATION NUMBER: US/08/316,714
Cc FILING DATE:

Cc CLASSIFICATION: 435
Ck PRIOR APPLICATION DATA:
Cc APPLICATION NUMBER: US/07/961,404
Cc FILING DATE: 15-OCT-1992
Cc ATTORNEY/AGENT INFORMATION:
Cc NAME: Goldman, Kenneth M.
Cc REGISTRATION NUMBER: 34,174
Cc REFERENCE/DOCKET NUMBER: 2595.1
Cc TELECOMMUNICATION INFORMATION:
Cc TELEPHONE: (510) 420-3152
Cc TELEFAX: (510) 658-5470
Cc TELEX: N/A
Cc INFORMATION FOR SEQ ID NO: 10:
Cc SEQUENCE CHARACTERISTICS:
Cc LENGTH: 585 amino acids
Cc TYPE: amino acid
Cc STRANDEDNESS: single
Cc TOPOLOGY: linear
Cc MOLECULE TYPE: protein
Cc SEQUENCE 585 AA; 65330 MW; 1874129 CN;

Query Match 28.7%; Score 219; DB 2; Length 585;
Best Local Similarity 35.9%; Pred. No. 6.96e-13;

[illegible]

OY 117 VALRGRKNEFEETVEYVYASVWGTVCSSHHDDSDASVICHQLOLGKGLA-KOTPFSGLG 175
Db 83 SCPIMDEVQCTGTEASLADC-KSLGMLKSNCRHERDAGVCT 124
OY 176 LPIYWSNVRCRGDEENILLCEKDI-WGGVCPQKMAAAVTC 217

RESULT 11
ID US-08-473-791-10 STANDARD; PRT; 585 AA.

AC xxxxxx

Sequence 10, Application US/08473791

DE Patent No. 5736340
XX Sequence 10, Application US/08473791

GENERAL INFORMATION:

APPLICANT: Kolts, Kirston E.

APPLICANT: Halenbeck, Robert F.

APPLICANT: Taylor, Eric W.

APPLICANT: Wang, Alice M.

APPLICANT: Caspitt, Clayton L.

TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cetus Oncology Corporation

STREET: 1400 Fifty-Third Street

CITY: Emeryville

STATE: CA

COUNTRY: USA

ZIP: 94608

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/473,791

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/316,714

FILING DATE:

APPLICATION NUMBER: US/07/961,404

FILING DATE: 15-OCT-1992

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Kenneth M.

REGISTRATION NUMBER: 34,174

REFERENCE/DOCKET NUMBER: 2595.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 420-3152

TELEFAX: (510) 658-5470

TELEX: N/A

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 585 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 585 AA: 65330 MW, 1874129 CN;

Query Match 28.7%; Score 219; DB 1; Length 585;

Best Local Similarity 35.9%; Pred. No. 6,96e-13;

Matches 37; Conservative 21; Mismatches 41; Indels 4; Gaps 4;

Db 24 MRLAGATNCRVEIFEFYWGOWGTCDNLDLTLDASVCRALGFENATGALGRAAF-GGG 82

OY 117 VALRGRKNEFEETVEYVYASVWGTVCSSHHDDSDASVICHQLOLGKGLA-KOTPFSGLG 175

Db 83 SCPIMDEVQCTGTEASLADC-KSLGMLKSNCRHERDAGVCT 124

OY 176 LPIYWSNVRCRGDEENILLCEKDI-WGGVCPQKMAAAVTC 217

RESULT 12
ID US-08-794-795-2 STANDARD; PRT; 495 AA.

AC xxxxxx

Sequence 2, Application US/08794795

DE Patent No. 5916766
XX Sequence 2, Application US/08794795

GENERAL INFORMATION:

APPLICANT: Elshouraghy, Nabil

APPLICANT: Adamou, John

APPLICANT: Lysko, Paul

TITLE OF INVENTION: Human Macro Scavenger Rec

TITLE OF INVENTION: eptor

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/794,795

FILING DATE: 04-FEB-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: ATG50009P

FILING DATE: 22-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: ATG50009

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-4026

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 495 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 495 AA: 49764 MW, 1218428 CN;

Query Match 28.4%; Score 217; DB 2; Length 495;

Best Local Similarity 41.2%; Pred. No. 1.12e-12;

Matches 42; Conservative 15; Mismatches 38; Indels 7; Gaps 5;

Db 399 VRIYSSNR-GRAYVYSGTWGTCDEDMONSDAIVFCRMLGYS-KGRALYKVGAGTG- 454

OY 117 VALRGRKNEFEETVEYVYASVWGTVCSSHHDDSDASVICHQLOLGKGLA-KOTPFSGLG 176

Db 455 -OI-MLDNVQCRGTESTLWSCYKNSGWHDCSHEDAGVCS 494

OY 177 IPIYWSNVRCRGDEENILLCEKDI-WGGVCPQKMAAAVTC 217

RESULT 13
ID US-08-794-795-6 STANDARD: PRT: 520 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 6, Application US/08794795
CC
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourig, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Lysko, Paul
CC APPLICANT: Gross, Mitchell
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA: 52658 MW: 1340662 CN:
DB 424 VRIVGSSNR--GRAEYVISTGWTICDDDEMONSDAIVFCMLGYS-KGRALIVYGAGTG-479
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQLDGGKAKOTPFSGGL 176
DB 480 -OI-WLDNVOCRGTESTLMSCTKNSMGHDCSHEDAGVCS 519
DY 177 IPIYW-SNVRCRDEENILLCERKDIMOGVCPKMAAAVICS 217
RESULT 14
ID US-08-470-350B-2 STANDARD: PRT: 1290 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 2, Application US/08470350B
CC
CC Sequence 2, Application US/08470350B
CC Patent No. 5664126
CC GENERAL INFORMATION:
CC APPLICANT: Li, Xiao
CC APPLICANT: Snyder, Solomon H
CC TITLE OF INVENTION: Enderin: A Secreted von Ebner's Gland
CC TITLE OF INVENTION: Protein Associated with Taste Buds
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1001 G Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,350B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wolfe, Susan A
CC REGISTRATION NUMBER: 33,568
CC REFERENCE/DOCKET NUMBER: 01107.48790
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-508-9100
CC TELEFAX: 202-508-9299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1290 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1290 AA: 141879 MW: 9315616 CN:
DB 145 VRLVNGDRCRGREVELLYGSGWGTCDSDINDANVCRQLGCGWALSAPGSAQF-GOG 203
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQLDGGKAKOTPFSGGL 175
DB 204 SCSIVLADVACRGHEAVLWCSHRGWLSHNCGHQEDAGVICS 245
DY 176 IPIYWSNVRCRDEENILLCERKDIMOGVCPKMAAAVICS 217
Query Match 28.4%; Score 217; DB 2; Length 520;
Best Local Similarity 41.2%; Pred. No. 1,12e-12;
Matches 42; Conservative 15; Mismatches 38; Indels 7; Gaps 5;

DT
XX
DE
XX
Sequence 2, Application US/08470350B
CC
XX
Sequence 2, Application US/08470350B
CC Patent No. 5664126
CC GENERAL INFORMATION:
CC APPLICANT: Li, Xiao
CC APPLICANT: Snyder, Solomon H
CC TITLE OF INVENTION: Enderin: A Secreted von Ebner's Gland
CC TITLE OF INVENTION: Protein Associated with Taste Buds
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1001 G Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,350B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wolfe, Susan A
CC REGISTRATION NUMBER: 33,568
CC REFERENCE/DOCKET NUMBER: 01107.48790
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-508-9100
CC TELEFAX: 202-508-9299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1290 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1290 AA: 141879 MW: 9315616 CN:
DB 145 VRLVNGDRCRGREVELLYGSGWGTCDSDINDANVCRQLGCGWALSAPGSAQF-GOG 203
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQLDGGKAKOTPFSGGL 175
DB 204 SCSIVLADVACRGHEAVLWCSHRGWLSHNCGHQEDAGVICS 245
DY 176 IPIYWSNVRCRDEENILLCERKDIMOGVCPKMAAAVICS 217
Query Match 27.7%; Score 212; DB 1; Length 1290;
Best Local Similarity 37.3%; Pred. No. 3.63e-12;
Matches 38; Conservative 16; Mismatches 46; Indels 2; Gaps 2;
DB 145 VRLVNGDRCRGREVELLYGSGWGTCDSDINDANVCRQLGCGWALSAPGSAQF-GOG 203
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQLDGGKAKOTPFSGGL 175
DB 204 SCSIVLADVACRGHEAVLWCSHRGWLSHNCGHQEDAGVICS 245
DY 176 IPIYWSNVRCRDEENILLCERKDIMOGVCPKMAAAVICS 217
RESULT 15
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 2, Application US/08200900A
CC
CC Sequence 2, Application US/08200900A
CC Patent No. 5665566
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

Thu Mar 16 07:57:48 2000

US-09-147-947-6-03.ra1

Page 8

```

CC STREET: 87 Cambridgepark Drive
CC City: Cambridge
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02140
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/200,900A
CC FILING DATE: 23-FEB-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meinerdt, Maureen C.
CC REGISTRATION NUMBER: 31,544
CC REFERENCE/DOCKET NUMBER: GI 5201-FWC
C8 TELECOMMUNICATION INFORMATION:
C8 TELEPHONE: (617) 876-1170 X8574
C8 TELEFAX: (617) 876-5851
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 798 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 798 AA; 88937 MW; 3292434 CN;
Query Match 14.8%; Score 113; DB 1; Length 798;
Best Local Similarity 28.6%; Pred. No. 1,936-02;
Matches 16; Conservative 10; Mismatches 29; Indels 1; Gaps 1;
D6 457 VRLNGTITDSGGLVQFRIOIIMHVAACAEWMTQTISDDYCOLLIGL-TGNSSVPFES 511
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
117 VRLKGGKMEFGICIVEIVASGVWGTVCSHWDDSDPVSVCHELQLGKGKIAOTPEFS 172

```

Search completed: Mon Mar 13 10:26:11 2000
Job time : 8 secs.

117-217/5KR1

use 117 2 43

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Run on: Mon Mar 13 10:24:02 2000; Maspar time 6.39 Seconds
 Tabular output not generated. 472.013 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (117-217) from US09147947A.ppt (3 of 6)
 Perfect Score: 764
 Sequence: 1 VRLRGKNEFEGETVEVYASG.....KDIWGGVCPQKMAAVTCS 101

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: Swiss-Prot38
 1:swissprot

Statistics: Mean 38.635; Variance 55.528; scale 0.696

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
1	764	100.0	875	1 NTR_HUMAN	1.53e-182
2	333	42.3	761	1 NTR_MOUSE	6.23e-19
3	266	33.5	454	1 MSRE_RABIT	2.76e-41
4	244	31.9	458	1 MSRE_MOUSE	3.40e-38
5	237	31.0	451	1 MSRE_HUMAN	2.10e-36
6	236	30.9	1436	1 WC11_BOVIN	3.78e-36
7	233	30.5	453	1 MSRE_BOVIN	2.19e-35
8	193	25.3	532	1 SPER_STRPU	2.13e-25
9	168	24.6	468	1 CD6_HUMAN	3.52e-24
10	114	14.9	1069	1 ENTK_MOUSE	2.27e-07
11	113	14.8	1035	1 ENTK_BOVIN	3.64e-07
12	107	14.0	1034	1 ENTK_PIG	5.99e-06
13	105	13.7	1019	1 ENTK_HUMAN	1.48e-05
14	102	13.4	583	1 CFAT_HUMAN	5.71e-05
15	98	12.8	495	1 CD5_BOVIN	3.35e-04
16	91	11.9	494	1 CD5_MOUSE	6.71e-03
17	90	11.8	326	1 PEL_EMENI	1.02e-02
18	88	11.5	3110	1 LAM2_HUMAN	2.32e-02
19	87	11.4	491	1 HMG2_YEAST	3.49e-02
20	84	11.0	495	1 CD5_HUMAN	1.16e-01
21	83	10.9	558	1 TF65_CHICK	1.77e-01
22	81	10.6	1857	1 FAS2_PENPA	3.76e-01
23	80	10.5	204	1 LEF2_NEVOP	5.52e-01

RESULT	ID	1	STANDARD	PRT	875 AA
AC	15-DEC-1999	(Rel. 39, Created)			
DT	15-DEC-1999	(Rel. 39, Last sequence update)			
DR	15-DEC-1999	(Rel. 39, Last annotation update)			
DE	NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).				
GN	PRSS12				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
NC	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-BRAIN:				
RX	MEDLINE: 98201705.				
RA	PROBA K., GSCHWEND T.P., SONDEREGGER P.,				
RT	"Cloning and sequencing of the cDNA encoding human neutrotypin."				
RL	Biochim. Biophys. Acta 1396:143-147(1998).				
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE TRYPSIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.				
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL: AJ001531; CAA04816.1; -				
DR	PROSITE: PS00134; TRYPSIN_HIS. 1.				
DR	PROSITE: PS00135; TRYPSIN_SER. 1.				
DR	PROSITE: PS00420; SPERACT_RECEPTOR. 3.				
KW	Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.				
FT	SIGNAL	1	20	POTENTIAL.	
FT	CHAIN	21	875	NEUTROTYPIN.	
FT	DOMAIN	23	92	PROLINE-RICH.	
FT	DOMAIN	93	165	KRINGLE.	
FT	DOMAIN	170	271	SRCR 1.	
FT	DOMAIN	170	280	SRCR 2.	
FT	DOMAIN	387	487	SRCR 3.	

HEX	ADEB3	HEXON PROTEIN (LATE PR	5.52e-01
24	80	10.5	911
25	79	10.3	446
26	78	10.2	491
27	78	10.2	576
28	77	10.1	186
29	77	10.1	186
30	77	10.1	1297
31	76	9.9	77
32	76	9.9	164
33	76	9.9	289
34	76	9.9	303
35	76	9.9	308
36	76	9.9	6359
37	75	9.8	327
38	75	9.8	458
39	75	9.8	1257
40	75	9.8	1332
41	74	9.7	61
42	74	9.7	168
43	74	9.7	338
44	74	9.7	340
45	74	9.7	340

```

FT DOMAIN 500 601 SRCR 4.
FT DOMAIN 619 875 SERINE PROTEASE.
FT DOMAIN 619 930 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 626 626 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 67D5272B CRC32:

Query Match 100.0%; Score 764; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.53e-182;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dn 170 VRLGSGKNEFGTEVYVAGWGTVCSSHMDDSDASVICHOLGKGIAKOTPFSGGL 229
117 VRLGSGKNEFGTEVYVAGWGTVCSSHMDDSDASVICHOLGKGIAKOTPFSGGL 176
230 IPIYWSNVRGDEENILLCERDIWOGVCPOKMAAAVTC 270
177 IPIYWSNVRGDEENILLCERDIWOGVCPOKMAAAVTC 217

RESULT 2
ID NETR_MOUSE STANDARD: PRT: 761 AA.
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN) (BRAIN-SPECIFIC SERINE
PROTEASE 3) (BSSP-3).
GN PRS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE: 97401523.
RA GSCHEMEND T.P., KRUGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neutrotypin, a novel multidomain serine protease expressed in the
RL nervous system.";
RN Mol. Cell. Neurosci. 9:207-219(1997).
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSURUOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RL "Molecular cloning of a novel brain-specific serine protease with a
triple-like structure and three scavenger receptor cysteine-rich
motifs.";
RN Biochem. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
AND AMYGDALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

DR EMBL: Y13192; CAA73646.1; -.
DR EMBL: D89871; BAA23986.1; -.
DR MGD: MGI:1100881; PRSS12.
DR PFAM: PF00530; SRCR: 3.
DR PFAM: PF00089; trypsin: 1.
DR PROSITE: PS00134; TRYPSIN_HIS: 1.
DR PROSITE: PS00135; TRYPSIN_SER: 1.
DR PROSITE: PS00420; SPERACT_RECEPTOR: 3.
DR Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 761 NEUTROTYPIN.
FT DOMAIN 85 157 KRINGLE.
FT DOMAIN 166 267 SRCR 1.
FT DOMAIN 273 373 SRCR 2.
FT DOMAIN 386 487 SRCR 3.
FT DOMAIN 505 761 SERINE PROTEASE.
FT DOMAIN 505 516 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 516 517 CHARGE RELAY SYSTEM.
FT ACT_SITE 562 562 CHARGE RELAY SYSTEM.
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 POTENTIAL.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA: 3F3C4F35 CRC32:

Query Match 42.3%; Score 323; DB 1; Length 761;
Best Local Similarity 45.5%; Pred. No. 6.23e-59;
Matches 46; Conservative 19; Mismatches 34; Indels 2; Gaps 2;

Dn 166 IRLVGSNGHEGRVLYHYAGWGTCIDQDMADADYICRQLGSGIAKAWHQAHF-GE 224
117 VRLGSGKNEFGTEVYVAGWGTVCSSHMDDSDASVICHOLGKGIA-KOTPFSGGL 175
225 SGPILLDEVRCTGNELSTFQCPKSSWGCHNGKEDAGVSC 265
176 IPIYWSNVRGDEENILLCERDIWOGVCPOKMAAAVTC 216

RESULT 3
ID MSRE_RABBIT STANDARD: PRT: 454 AA.
AC 005585;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
LDL RECEPTOR I AND II).
GN MSRI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93016877.
RA BICKEL P.E., FREEMAN M.W.;
RT "Rabbit aortic smooth muscle cells express inducible macrophage
scavenger receptor messenger RNA that is absent from endothelial
cells.";
RN J. Clin. Invest. 90:1450-1457(1992).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93131972.
RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
RA MIYAZAKI T., NAKAMURA H., UESUGI S., IWANISHI T., KAWABE Y.,
RA IMAKURA H., YAZAKI Y., MATSUMOTO A.;
RT "Charged collagen structure mediates the recognition of negatively
charged macromolecules by macrophage scavenger receptors.";
RL J. Biol. Chem. 268:2126-2133(1993).
CC -1- FUNCTION: MEMBRANE GLYCOPROTEIN IMPLICATED IN THE PATHOLOGIC
DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE
ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING

```

CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L11693; AAA31402.1; -;
 CC EMBL: L11692; AAA31403.1; -;
 CC EMBL: D13381; BAA02649.1; -;
 CC PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC
 CC DR PFAM: PF00530; SRCR; 1.
 CC DR Transmembrane; Glycoprotein; 1.
 CC KM Heptad repeat pattern: Receptor: Alternative splicing.
 CC FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 51 73 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC FT DOMAIN 74 454 (POTENTIAL).
 CC FT DOMAIN 74 109 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 110 272 SPACER (PROBABLE).
 CC FT DOMAIN 273 344 ALPHA-HELICAL COILED-COIL.
 CC FT DOMAIN 345 434 COLLAGEN-LIKE.
 CC FT DOMAIN 82 82 CIS-RICH.
 CC FT CARBOHYD 102 102 POTENTIAL.
 CC FT CARBOHYD 143 143 POTENTIAL.
 CC FT CARBOHYD 184 184 POTENTIAL.
 CC FT CARBOHYD 221 221 POTENTIAL.
 CC FT CARBOHYD 249 249 POTENTIAL.
 CC FT CARBOHYD 267 267 POTENTIAL.
 CC FT VARSPLIC 348 354 TPSAIR -> REVOLTP (IN ISOFORM II).
 CC FT VARSPLIC 355 454 MISSING (IN ISOFORM II).
 CC FT CONFLICT 106 106 H -> D (IN REF. 2).
 CC SQ SEQUENCE 454 AA: 49745 MW: 5D780348 CRC32:
 CC
 CC Query Match 33.5%; Score 256; DB 1; Length 454;
 CC Best Local Similarity 40.8%; Pred. No. 2,76e-41;
 CC Matches 42; Conservative 17; Mismatches 40; Indels 4; Gaps 4;
 CC
 CC 353 VRLVGRGPHREGVEILHNGOMGTCDHWEHLRAGOVCSLGIRGVKSHKAYF-GQG 411
 CC 117 VRLRGGRNEDEGEVEYVAVSGWGTVCSSHWDDSDASVICHQLDL-GGKGIKOTPFESGLG 175
 CC DB 412 TGPI-WLNEVPCLGMESSECKIRQGVAVCSHGEDAGYTC 453
 CC QY 176 LPIIYW-SNVRCRGDEENILLCEKDIWGGVCPQKMAAAVTC 217
 CC
 CC RESULT 4
 CC ID MSRE.MOUSE STANDARD; PRT; 458 AA.
 CC AC P30204;
 CC DT 01-APR-1993 (Rel. 25, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-DEC-1999 (Rel. 39, Last annotation update)
 CC DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 CC LDL RECEPTOR I AND II).
 CC GN MSRL OR SCVR.
 CC OS Mus musculus (mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 93359822.
 CC RA ASHKENAS J., PENMAN M., VASILE E., ACTON S., FREEMAN M.W.,
 CC KRIEGER M.;

RT "Structures and high and low affinity ligand binding properties of
 RT murine type I and type II macrophage scavenger receptors.";
 RT J. Lipid Res. 34:983-1000(1993).
 RL [2]
 RN RP SEQUENCE FROM N.A. (SHORT FORM).
 RX MEDLINE: 93331972.
 RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
 RA MIYAKAKI T., NAKAMURA H., DESUO S., IMANISHI T., KAMAE Y.,
 RA ITAKURA H., YAZAKI Y., MATSUMOTO A.;
 RT "Charged collagen structure mediates the recognition of negatively
 RT charged macromolecules by macrophage scavenger receptors.";
 RL J. Biol. Chem. 268:2126-2133(1993).
 RN [3]
 RP SEQUENCE OF 349-458 FROM N.A.
 RX MEDLINE: 91062370.
 RA FREEMAN M., ASHKENAS J., REES D.J., KINGSLEY D.M., COPELAND N.G.,
 RA JENKINS N.A., KRIEGER M.;
 RT "An ancient, highly conserved family of cysteine-rich protein domains
 RT revealed by cloning type I and type II murine macrophage scavenger
 RT receptors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).
 RN [4]
 RP SEQUENCE OF 1-4 FROM N.A.
 RX MEDLINE: 95395388.
 RA AETRING R.P., FREEMAN M.W.;
 RT "Structure of the murine macrophage scavenger receptor gene and
 RT evaluation of sequences that regulate expression in the macrophage
 RT cell line, P388D.";
 RL J. Lipid Res. 36:1305-1314(1995).
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIANE THE
 CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: I04274; AAA39747.1; ALT_INIT.
 CC DR EMBL: I04275; AAA39748.1; ALT_INIT.
 CC DR EMBL: D13382; BAA02650.1; -;
 CC DR EMBL: M59445; BAA37464.1; -;
 CC DR EMBL: M59446; AAA37465.1; -;
 CC DR EMBL: U13873; AAC13774.1; -;
 CC DR PIR: A38260; A38260.
 CC DR MGI: 98257; SCVR.
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC DR PFAM: PF00530; SRCR; 1.
 CC DR Transmembrane; Glycoprotein; 1.
 CC KM Heptad repeat pattern: Receptor: Alternative splicing.
 CC FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 56 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC FT DOMAIN 79 458 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 79 114 SPACER (PROBABLE).
 CC FT DOMAIN 114 276 ALPHA-HELICAL COILED-COIL.
 CC FT DOMAIN 277 350 COLLAGEN-LIKE.
 CC FT DOMAIN 351 458 CIS-RICH.
 CC FT CARBOHYD 94 94 POTENTIAL.
 CC FT CARBOHYD 107 107 POTENTIAL.
 CC FT CARBOHYD 147 147 POTENTIAL.
 CC FT CARBOHYD 168 188 POTENTIAL.

FT CARBOHYD 253 253 POTENTIAL.
 FT CARBOHYD 271 271 POTENTIAL.
 FT VARSPLIC 352 354 TPL -> RSV (IN ISOFORM II).
 FT VARSPLIC 355 458 MISSING (IN ISOFORM II).
 SQ SEQUENCE 458 AA; 50130 MW; F28A456E CRC32;

Query Match
 Best Local Similarity 36.9%; Score 244; DB 1; Length 458;
 Matches 38; Conservative 19; Mismatches 42; Indels 4; Gaps 4;

DB 357 VRLVGGSGAEGREYVEIHOGOMGTICDDRMIDRAGQVCRSLGYOEVLAVHRAHF-GOG 415
 117 VRLRGKNEEGVEYVYASGVWGVCSSHMDSDASYICHQLDGLG-KGIAKOTPFSGLG 175

DB 416 TGP1-WLNEVYCFGRSSIECKIRMGVLSCHSHSDEGAYTCT 457
 176 LIPIYW-SNVRCRDEENILLCERDIWOGVCPOKMAAVTCS 217

5
 ASRE-HUMAN STANDARD; PRT; 451 AA.
 P2157; P21759;
 01-MAY-1991 (Rel. 18, Created)
 01-MAY-1991 (Rel. 18, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 LDL RECEPTOR I AND II).
 GN MSRI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE; 91067661.
 RA MATSUMOTO A., NAITO M., ITAKURA H., IKEMOTO S., ASAKA H.,
 HAYAKAWA I., ABURATANI H., TAKAKU F., SUZUKI H.,
 KOBARI Y., MIYAI T., TAKAHASHI K., COHEN E.H., WIDRO R.,
 HOSHMAN D.E., KODAMA T.;
 "Human macrophage scavenger receptors: primary structure, expression,
 and localization in atherosclerotic lesions."
 Proc. Natl. Acad. Sci. U.S.A. 87:9133-9137(1990).
 CC -! FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
 ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -! SUBUNIT: HOMOTRIMER.
 CC -! SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -! ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 OF THE SAME GENE.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90187; BAA14208.1; -
 DR EMBL: D90188; BAA14209.1; -
 DR PIR: A38415; A38415.
 DR PIR: B38415; B38415.
 DR MIM: 153622; -
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF01391; Collagen; 1.
 KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 Heptad repeat pattern; Receptor; Alternative splicing;
 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1 50 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT TRANSMEM 51 76
 FT DOMAIN 77 451 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 77 109 SPACER (PROBABLE).
 FT DOMAIN 110 272 ALPHA-HELICAL COILED-COIL.
 FT DOMAIN 273 341 COLLAGEN-LIKE.
 FT DOMAIN 342 451 CYS-RICH.
 FT CARBOHYD 82 82 POTENTIAL.
 FT CARBOHYD 102 102 POTENTIAL.
 FT CARBOHYD 143 143 POTENTIAL.
 FT CARBOHYD 184 184 POTENTIAL.
 FT CARBOHYD 221 221 POTENTIAL.
 FT CARBOHYD 249 249 POTENTIAL.
 FT CARBOHYD 267 267 POTENTIAL.
 FT VARSPLIC 345 358 POTENTIAL.
 FT VARSPLIC 359 451 ITPKTRIVGSGP -> RPVLDTIRAGPS (IN
 ISOFORM II).
 SQ SEQUENCE 451 AA; 49762 MW; AB9FECE7 CRC32;

Query Match
 Best Local Similarity 37.9%; Score 237; DB 1; Length 451;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPEGRVEIILHSGONGTICDDREVEVGVCRSLGYPGOVHRAHF-GOG 408
 117 VRLRGKNEEGVEYVYASGVWGVCSSHMDSDASYICHQLDGLG-KGIAKOTPFSGLG 175

DB 409 TGP1-WLNEVYCFGRSSIECKIRMGVLSCHSHSDEGAYTCT 450
 176 LIPIYW-SNVRCRDEENILLCERDIWOGVCPOKMAAVTCS 217

RESULT 6
 ID WC11_BOVIN STANDARD; PRT; 1436 AA.
 AC P30205;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE ANTIGEN WC1.1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BLOOD.
 RX MEDLINE; 93056489.
 RA WJUNGARD P.L.J., METZELAR M.J., MACHUGH N.D., MORRISON W.I.,
 CLEVERS H.C.;
 RT "Molecular characterization of the WC1 antigen expressed specifically
 on bovine CD4-CD8-gamma delta T lymphocytes."
 RL J. Immunol. 149:3273-3277(1992).
 CC -----
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X63723; CAA5255.1; -
 DR PIR: S19913; S19913.
 DR PIR: A46496; A46496.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 4.
 DR PFAM: PF00530; SRCR; 1.
 KW Antigen; Repeat.
 SQ SEQUENCE 1436 AA; 154196 MW; D61545D7 CRC32;

Query Match
 Best Local Similarity 30.9%; Score 236; DB 1; Length 1436;
 Matches 39; Conservative 20; Mismatches 40; Indels 4; Gaps 4;

DB 1155 LRLRGSGSESGREYVHNSMGTCVDDSDSLAEAVVCOGLGCGALAVSAAP-GPG 1213
 117 VRLRGKNEEGVEYVYASGVWGVCSSHMDSDASYICHQLDGLG-KGIAKOTPFSGLG 175

Db 1214 NSST-WLDEVQCGGREGSSIMDCVAEPWGOSDCKHEPDAGRCS 1255
 176 LPIIYW-SNVRGDEENILCEKIDIMOGVCPQKMAAVTCS 217

RESULT 7 STANDARD: PRT: 453 AA.

AC P21758;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 LDL RECEPTOR I AND II).
 CS Bos taurus (Bovine).
 CS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Outeria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 Bovinae; Bos.

[1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE-LUNG;
 RX MEDLINE: 90136965.
 RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
 RA KRIEGER M.,
 RT "Type I macrophage scavenger receptor contains alpha-helical and
 RT collagen-like coiled coils.";
 RL Nature 343:531-535(1990).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).

RC TISSUE-LUNG;
 RX MEDLINE: 90136973.
 RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIEGER M.,
 RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
 RT scavenger receptor type II.";
 RL Nature 343:570-572(1990).

CC -I- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIANE THE
 CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).

CC -I- SUBUNIT: HOMOTRIMER.
 CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -I- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X51689; CA35987.1;
 DR EMBL: X54183; CA38108.1;
 DR PIR: S08276; S08276.
 DR PIR: S08278; S08278.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF01391; Collagen; 1.
 KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 KW Heptad repeat pattern; Receptor; Alternative splicing.
 FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 51 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 77 453 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 77 108 SPACER (PROBABLE).
 FT DOMAIN 109 271 ALPHA-HELICAL COILED-COIL.
 FT DOMAIN 272 343 COLLAGEN-LIKE.
 FT DOMAIN 344 453 CYS-RICH.
 FT CARBOHYD 82 82 POTENTIAL.
 FT CARBOHYD 101 101 POTENTIAL.
 FT CARBOHYD 142 142 POTENTIAL.

FT CARBOHYD 183 183 POTENTIAL.
 FT CARBOHYD 220 220 POTENTIAL.
 FT CARBOHYD 248 248 POTENTIAL.
 FT CARBOHYD 266 266 POTENTIAL.
 FT VARSPLIC 348 349 OS -> PG (IN ISOFORM II).
 FT VARSPLIC 350 453 MISSING (IN ISOFORM II).
 SQ SEQUENCE 453 AA; 50056 MW; 2CDEIEEL CRC32;

Query Match 30.5%; Score 233; DB 1; Length 453;
 Best Local Similarity 37.9%; Pred. No. 2,19e-35;
 Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

Db 352 VRLVGGSGPHEGRVEIFHEGQNGTVCDDWELRGILVCRSLGYGVQSVKRAYF-GKG 410
 117 VRLRGKNEFEESTVEYASGVGTCCSSHWDDSDASVICHQLQ-GGKGIARQTPSSGLG 175
 Db 411 TGPI-WLNVFPGKRESSIEERIRQWGRACSHDEDAVTC 452
 176 LPIIYW-SNVRGDEENILCEKIDIMOGVCPQKMAAVTCS 217

RESULT 8 STANDARD: PRT: 532 AA.

AC P16264;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EGG PEPTIDE SPERACT RECEPTOR PRECURSOR.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 OC Echinoidea; Echinoidea; Echinoidea; Strongylocentrotidae;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 477-489.
 RX MEDLINE: 89184581.

RA DANCOTT L.J., JORDAN J.E., BELLET R.A., GARBERS D.L.;
 RT "Cloning of the mRNA for the protein that crosslinks to the egg
 RT peptide speract.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).
 CC -I- FUNCTION: RECEPTOR FOR THE EGG PEPTIDE SPERACT.

CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: J04518; AAA30078.1;
 DR PIR: A32751; A32751.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 4.
 DR PFAM: PF00530; SRCR; 4.
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
 FT SIGNAL 1 30
 FT CHAIN 31 532 EGG PEPTIDE SPERACT RECEPTOR.
 FT DOMAIN 31 491 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 492 520 POTENTIAL.
 FT DOMAIN 521 532 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 40 488 4 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 40 147 1.
 FT REPEAT 148 260 2.
 FT REPEAT 261 375 3.
 FT REPEAT 376 488 4.
 FT CARBOHYD 78 78 POTENTIAL.
 FT CARBOHYD 115 115 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 SQ SEQUENCE 532 AA; 57820 MW; D2174AE CRC32;

Query Match 25.3%; Score 193; DB 1; Length 532;
 Best Local Similarity 35.8%; Pred. No. 2,13e-25;
 Matches 38; Conservative 19; Mismatches 41; Indels 8; Gaps 6;

Db 264 IRMDGSPHREVRVIMHDANGTICDDGMDANAVNCRQ-A-GYRGVAKSGFKGEDE 321
 QY 117 VRLRGKNEFEETVEYVAGVGTWCSSHMDDSDASVICHQLOLGKGIKQKPFSG--L 174
 Db 322 GFTWAPIHTEFVCTGVEDRLIDCIRLDGWTMS-CYHVDASVCA 366
 QY 175 GLI--PIYMSNVRCRGDEENILCE-KDIMGQVCPQKMAAVTCS 217

RESULT 9
 ID CD6_HUMAN STANDARD: PRT: 468 AA.
 AC P30203:
 DI 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL DIFFERENTIATION ANTIGEN CD6 PRECURSOR (T12) (T120).
 GN CD6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 RN Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92013796.
 RA ARUFFO A., MELNICK M.B., LINSLEY P.S., SEED B.:
 RT "The lymphocyte glycoprotein CD6 contains a repeated domain structure
 characteristic of a new family of cell surface and secreted
 proteins.";
 RT J. Exp. Med. 174:949-952(1991).
 CC -1- FUNCTION: INVOLVED IN CELL ADHESION. BINDS TO CD166.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD6 entry;
 WWW="http://www.ncbi.nlm.nih.gov/PROV/cd/cd6.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X60992; CAA43306.1; -
 DR PIR: S26741; S26741.
 DR MIM: 186720; -
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 DR PFAM: PF00530; SRCR; 3.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal.
 FT SIGNAL 1
 FT MAIN ? 468 T-CELL DIFFERENTIATION ANTIGEN CD6.
 FT DOMAIN ? 402 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 403 423 POTENTIAL.
 FT DOMAIN 424 468 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 28 28 POTENTIAL.
 FT CARBOHYD 49 49 POTENTIAL.
 FT CARBOHYD 112 112 POTENTIAL.
 FT CARBOHYD 118 118 POTENTIAL.
 FT CARBOHYD 229 229 POTENTIAL.
 FT CARBOHYD 339 339 POTENTIAL.
 FT CARBOHYD 345 345 POTENTIAL.
 FT CARBOHYD 368 368 POTENTIAL.
 SQ SEQUENCE 468 AA; 49828 MW; 042D5AC8 CRC32;

Query Match 24.6%; Score 188; DB 1; Length 468;
 Best Local Similarity 36.6%; Pred. No. 3 52e-24;
 Matches 37; Conservative 19; Mismatches 33; Indels 6; Gaps 4;

Db 266 RLITGADRCGQVEVHFRVNTVCDSEWYPSAEVLCQSLG-CGTAVERPGLPHSIS- 323
 QY 118 RLRGKNEFEETVEYVAGVGTWCSSHMDDSDASVICHQLOLGKGIKQKPFSG--L 176
 Db 324 GRWYYS---CNGEELTJLSCMRFNNSNLCSSSLAARVICS 361

QY 177 IPTYMSNVRCRGDEENILCEKDIMGQVCPQKMAAVTCS 217
 RESULT 10
 ID ENTK_MOUSE STANDARD: PRT: 1069 AA.
 AC P97435:
 DI 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 RN Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=DUODENUM;
 RA YUAN X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.:
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PIV: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U73378; AA37317.1; -
 DR HSBP; P00763; IDPO.
 DR MGD; MGI:1197523; PRSS7.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; TRYPSIN_2; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS01209; IDIRA_1; 2.
 DR PROSITE: PS50068; IDIRA_2; 2.
 DR PFAM: PF00057; Idl_recept_a; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00431; CUB; 2.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF00629; MAM; 1.
 DR PFAM: PF01390; SEA; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 FT Setine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 270 379 CUB.
 FT DOMAIN 387 549 MAM.
 FT DOMAIN 569 679 CUB.

FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 723 816 SRCR.
 FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1021 1021 MYRISTATE (POTENTIAL).
 FT LIPID 2 2
 FT DISULFID 229 242 BY SIMILARITY.
 FT DISULFID 236 255 BY SIMILARITY.
 FT DISULFID 249 266 BY SIMILARITY.
 FT DISULFID 688 700 BY SIMILARITY.
 FT DISULFID 695 713 BY SIMILARITY.
 FT DISULFID 707 722 BY SIMILARITY.
 FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 859 875 BY SIMILARITY.
 FT DISULFID 959 1027 BY SIMILARITY.
 FT DISULFID 991 1006 BY SIMILARITY.
 FT DISULFID 1017 1045 BY SIMILARITY.
 FT CARBOHYD 147 147 POTENTIAL.
 FT CARBOHYD 197 197 POTENTIAL.
 FT CARBOHYD 212 212 POTENTIAL.
 FT CARBOHYD 373 373 POTENTIAL.
 FT CARBOHYD 380 380 POTENTIAL.
 FT CARBOHYD 433 433 POTENTIAL.
 FT CARBOHYD 515 515 POTENTIAL.
 FT CARBOHYD 579 579 POTENTIAL.
 FT CARBOHYD 675 675 POTENTIAL.
 FT CARBOHYD 727 727 POTENTIAL.
 FT CARBOHYD 751 751 POTENTIAL.
 FT CARBOHYD 770 770 POTENTIAL.
 FT CARBOHYD 791 791 POTENTIAL.
 FT CARBOHYD 897 897 POTENTIAL.
 FT CARBOHYD 936 936 POTENTIAL.
 FT CARBOHYD 999 999 POTENTIAL.
 SO SEQUENCE 1069 AA; 118735 MM; 48B825A2 CRC32;

Query Match 14.9%; Score 114; DB 1; Length 1069;
 Best Local Similarity 26.7%; Pred. No. 2.27e-07;
 Matches 16; Conservative 17; Mismatches 25; Indels 2; Gaps 2;

Db 723 VRFNGTRSN-NGIYOVNINISIMHACENMTQISNEVCHLGLGSANSMPISSTGGG 781
 117 VR-LRGKNEFEGETVEYASGWMGTVCSSHWSDASVICHOLQLGKGIKOTPSGLG 175

RESULT 11 STANDARD: PRT: 1035 AA.
 ID ENTK_BOVIN 988072;

01-FEB-1996 (Rel. 33, Created)
 01-FEB-1996 (Rel. 33, Last sequence update)
 15-DEC-1999 (Rel. 39, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PPS7 OR ENTK.

OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 RN Bovinae; Bos.

RP [1]
 RC TISSUE-DUODENUM;
 RX MEDLINE: 94329561.

RA KILAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
 "Enterokinase, the initiator of intestinal digestion, is a mosaic
 protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).

RN [2]
 RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 94043122.
 RA LAVALLE E.R., REHEMUTULA A., RACIE L.A., DIBLASIO E.A.,
 FERENZ C., GRANT R.L., LIGHT A., MCCOY J.M.;
 "Cloning and functional expression of a cDNA encoding the catalytic
 subunit of bovine enterokinase.";
 RL J. Biol. Chem. 268:23311-23317(1993).
 RN [3]

RP SEQUENCE OF 801-827.
 RC TISSUE-INTESTINE;
 RX MEDLINE: 92189715.
 RA LIGHT A., JANSKA H.;
 RT "The amino-terminal sequence of the catalytic subunit of bovine
 enterokinase.";
 RL J. Protein Chem. 10:475-480(1991).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: U09859; AAB40026.1;
 CC EMBL: L19663; AAL16035.1;
 CC PIR: A61436; A61436.
 CC HSP: P00763; IDPO.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC PROSITE: PS01180; CUB; 2.
 CC PROSITE: PS00740; MAM_1; 1.
 CC PROSITE: PS50060; MAM_2; 1.
 CC PROSITE: PS01209; LDLRA_1; 2.
 CC PROSITE: PS50068; LDLRA_2; 2.
 CC PFAM: PF00057; ldl_recept_a; 2.
 CC PFAM: PF00089; trypsin; 1.
 CC PFAM: PF00431; CUB; 2.
 CC PFAM: PF00530; SRCR; 1.
 CC PFAM: PF00629; MAM; 1.
 CC PFAM: PF01390; SEA; 1.
 CC Signal anchor: Glycoprotein: Myristate; Hydrolyase;
 CC Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
 CC CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 CC CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
 CC TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 CC DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
 CC DOMAIN 240 350 CUB.
 CC DOMAIN 358 520 MAM.
 CC DOMAIN 520 650 CUB.
 CC DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
 CC DOMAIN 694 787 SRCR.
 CC ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 882 882 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 987 987 MYRISTATE (POTENTIAL).
 CC LIPID 2 2
 CC DISULFID 199 212 BY SIMILARITY.
 CC DISULFID 206 225 BY SIMILARITY.
 CC DISULFID 219 236 BY SIMILARITY.
 CC DISULFID 659 671 BY SIMILARITY.
 CC DISULFID 666 684 BY SIMILARITY.
 CC DISULFID 678 693 BY SIMILARITY.

FT CARBOHYD 804 804 POTENTIAL.
 FT CARBOHYD 863 863 POTENTIAL.
 FT CARBOHYD 902 902 POTENTIAL.
 FT CARBOHYD 964 964 POTENTIAL.
 SO SEQUENCE 1034 AA; 114776 MW; 24386471 CRC32;

Query Match 14.0%; Score 107; DB 1; Length 1034;
 Best Local Similarity 28.6%; Pred. No. 5,93e-06;
 Matches 17; Conservative 12; Mismatches 29; Indels 1; Gaps 1;

Db 693 VREFNGTANNGLVORFRIOSIMHTACAEWNTQISDVCCOLLIG-TGNSMPEFSSGG 750
 117 VRLRGKNEFEIVYASGVWGTVCSSHWDDSDASYICHQLQGGKGIKOTPPFSLG 175

Query Match 13
 ENTK_HUMAN STANDARD; PRT; 1019 AA.
 98073:
 01-FEB-1996 (Rel. 33, Created)
 01-FEB-1996 (Rel. 33, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=DUODENUM;
 RX MEDLINE; 95234679.
 RA KITAMOTO Y., VEILE R.A., DONIS-KELLER H., SADLER J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 PI the proteolytic activator of trypsinogen.";
 PL Biochemistry 34:4562-4568(1995).
 [2]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=DUODENUM;
 RX MEDLINE; 94329561.
 KA KITAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 PI protease composed of a distinctive assortment of domains.";
 PL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES.
 - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
 TRYPSINOGEN.
 - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 - TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 - DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
 MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
 - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 - SIMILARITY: CONTAINS 2 CUB DOMAINS.
 - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; 009860; AAC50138.1;

DR HSP; P00763; IDPO.
 DR MIM; 226200; -.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00060; MAM_2; 1.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PFAM; PF00057; LDL_recept_a; 2.
 DR PFAM; PF00089; trypsin; 1.
 DR PFAM; PF00431; CUB; 2.
 DR PFAM; PF00530; SRCR; 1.
 DR PFAM; PF00629; MAM; 1.
 DR PFAM; PF01390; SEA; 1.
 DR Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 KW Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 1 784
 FT CHAIN 1 784
 FT TRANSMEM 19 47
 FT DOMAIN 182 223
 FT DOMAIN 225 334
 FT DOMAIN 342 504
 FT DOMAIN 524 634
 FT DOMAIN 641 679
 FT DOMAIN 678 771
 FT ACT_SITE 825 825
 FT ACT_SITE 825 825
 FT ACT_SITE 876 876
 FT ACT_SITE 971 971
 FT LIPID 2
 FT DISULFID 184 197
 FT DISULFID 191 210
 FT DISULFID 204 221
 FT DISULFID 643 655
 FT DISULFID 650 668
 FT DISULFID 662 677
 FT DISULFID 772 896
 FT DISULFID 810 826
 FT DISULFID 910 977
 FT DISULFID 941 956
 FT DISULFID 967 995
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 179 179
 FT CARBOHYD 328 328
 FT CARBOHYD 335 335
 FT CARBOHYD 388 388
 FT CARBOHYD 440 440
 FT CARBOHYD 470 470
 FT CARBOHYD 503 503
 FT CARBOHYD 534 534
 FT CARBOHYD 630 630
 FT CARBOHYD 682 682
 FT CARBOHYD 706 706
 FT CARBOHYD 725 725
 FT CARBOHYD 848 848
 FT CARBOHYD 887 887
 FT CARBOHYD 909 909
 FT CARBOHYD 949 949
 SO SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;

Query Match 13.7%; Score 105; DB 1; Length 1019;
 Best Local Similarity 28.6%; Pred. No. 1.48e-05;
 Matches 16; Conservative 13; Mismatches 26; Indels 1; Gaps 1;

Db 678 VREFNGTNNGLVPRFRIOSIMHTACAEWNTQISDVCCOLLIGSGNSKPI-FS 732
 117 VRLRGKNEFEIVYASGVWGTVCSSHWDDSDASYICHQLQGGKGIKOTPPS 172

Query Match 14
 ENTK_HUMAN STANDARD; PRT; 583 AA.
 AC P05156;

13-AUG-1987 (Rel. 05, Created)
 13-AUG-1987 (Rel. 05, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 COMPLEMENT FACTOR I PRECURSOR (EC 3.4.21.45) (C3B/C4B INACTIVATOR).
 IF.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 [1]
 SEQUENCE FROM N.A.
 TISSUE=LIVER;
 MEDLINE: 87241401.
 RA CATERALL C.F., LYONS A., SIM R.M., DAY A.J., HARRIS T.J.R.;
 RT "Characterization of primary amino acid sequence of human complement
 RL Biochem. J. 242:849-856(1987).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE: 87280021.
 OLDENBERG G., BRUNS G.A.P., RITS M., EDGE M.D., KWATKOWSKI D.J.;
 RL Human complement factor I: analysis of cDNA-derived primary
 structure and assignment of its gene to chromosome 4.";
 J. Biol. Chem. 262:10065-10071(1987).
 CC -1- FUNCTION: FACTOR I IS RESPONSIBLE FOR CLEAVING THE ALPHA-CHAINS
 OF C4B AND C3B IN THE PRESENCE OF THE COFACTORS C4-BINDING
 PROTEIN AND FACTOR H RESPECTIVELY.
 CC -1- SUBUNIT: HETERODIMER OF A LIGHT AND HEAVY CHAINS LINKED BY
 DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPsin FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Y00318; CA68416.1; ALT_INIT.
 DR EMBL: J02770; AA52455.1; -.
 DR PIR: A29154; A29154.
 DR HSSP: P00763; IDPO.
 DR SWISS-2DPAGE: P05156; HUMAN.
 DR MIM: 217030; -.
 DR PROSITE: PS00134; TRYPSIN_HTS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS50068; LDLRA_2; 2.
 DR PFAM: PF00057; ldl_recept_a; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00530; SRCR; 1.
 KW Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
 KW Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 583
 FT CHAIN 19 583
 FT CHAIN 340 583
 FT DOMAIN 220 288
 FT DOMAIN 229 295
 FT DISULFID 229 247
 FT DISULFID 241 ?
 FT DISULFID 259 271
 FT DISULFID 266 284
 FT DISULFID 278 293
 FT DISULFID 365 381
 FT DISULFID 467 531
 FT DISULFID 495 510
 FT DISULFID 521 550
 FT ACT_SITE 380 380
 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 429 429 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 525 525 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 70 70 POTENTIAL.
 FT CARBOHYD 103 103 POTENTIAL.
 FT CARBOHYD 177 177 POTENTIAL.
 FT CARBOHYD 464 464 POTENTIAL.
 FT CARBOHYD 494 494 POTENTIAL.
 FT CARBOHYD 536 536 POTENTIAL.
 FT CONFLICT 558 558 V -> F (IN REF. 2).
 SQ SEQUENCE 583 AA; 65720 MW; D673E9F0 CRC32;
 Query Match 13.4%; Score 102; DB 1; Length 583;
 Best Local Similarity 28.9%; Pred. No. 5,71e-05;
 Matches 24; Conservative 18; Mismatches 38; Indels 3; Gaps 3;
 Db 114 VSUKHNTDSEGVIEKVLVDQDKTMEICKSSMSSEANVACDLGFGQADTQRRFKLSD 173
 117 VRLRGKNEFEQVEY-VASGWGT-VCSHDDSDASVICHQLDQ-GKGIATQTPPSG 173
 174 LGIPYMSVNCRGDENILLC 196
 QY
 Db 174 LSTNSTECLVHCRGLETSLAEC 196
 174 LGIPYMSVNCRGDENILLC 196
 QY
 RESULT 15
 ID CD5_BOVIN STANDARD; PRT; 495 AA.
 AC P19238;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCYTE GLYCOPROTEIN
 TI/LEU-1) (LYMPHOCYTE ANTIGEN CD5).
 GN CD5.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-THYMUS;
 RX MEDLINE: 90384847.
 RA YU Q., REICHERT M., BROUSSEAU T., CLEUTER Y., BURRY A., KETTMANN R.;
 RT "Sequence of bovine CD5.";
 RL Nucleic Acids Res. 18:5296-5296(1990).
 CC -1- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL
 CC PROLIFERATION. CD5 INTERACTS WITH CD72/LEB-2.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 SRCR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X53061; CA37231.1; -.
 DR PIR: S11270; S11270.
 DR PFAM: PF00530; SRCR; 3.
 KW Signal; Transmembrane; Glycoprotein; T-cell.
 FT SIGNAL 1 24
 FT CHAIN 25 495
 FT CHAIN 25 495
 FT CHAIN 372 495
 FT TRANSMEM 373 402
 FT DOMAIN 403 495
 FT DOMAIN 403 495
 FT DOMAIN 403 495
 FT DOMAIN 403 495
 FT DISULFID 43 108
 FT DISULFID 51 117
 FT CARBOHYD 117 117
 FT CARBOHYD 323 323
 FT CARBOHYD 369 369
 POTENTIAL.

SQ SEQUENCE 495 AA; 5433 MW; F8ED2AE0 CRC32;
 Query Match 12.8%; Score 98; DB 1; Length 495;
 Best Local Similarity 39.1%; Pred. NO. 3.35e-04;
 Matches 18; Conservative 9; Mismatches 15; Indels 4; Gaps 4;
 Db 275 RLVGSDVCEGSEVFRSGKGKMDLDDSWAKGTARREVCREQQ 320
 QY 118 RLRGKNEFEETVEYAS-GV-WGTVCSSHWDSDAS-V-ICHOLQ 159

Search completed: Mon Mar 13 10:24:13 2000
 Job time : 11 secs.

THIS PAGE BLANK (USPTO)

117-217/SCRCR1

No/Many before/Now PIRAS Saispact lists

Release 3.1a John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:24:29 2000; MasPar time 16.49 Seconds
Tabular output not generated. 424.656 Million cell updates/sec

Title: >US-09-147-947-6
Description: (117-217) from US09147947A.pep (3 of 6)
Perfect Score: 764
Sequence: 1 VRLRGKNEFEVYASG.....KDIMOGVCYCPQMAAAVTC 101

Scoring table: PAM 150
Gap 11
Searched: 225878 seqs, 69334122 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: 1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mbc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 37.758; Variance 56.548; scale 0.668
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	266	34.8	2153	5	097375	SCAVENGER RECEPTOR CYS	4.54e-42
2	254	33.2	2043	5	096943	SRCR DOMAIN, MEMBRANE	4.51e-39
3	251	32.9	1036	5	097378	SCAVENGER RECEPTOR CYS	2.51e-38
4	251	32.9	1116	4	007898	M130 ANTIGEN PRECURSOR	2.51e-38
5	251	32.9	1149	4	007901	M130 ANTIGEN, EXTENDED	2.51e-38
6	251	32.9	1151	4	007899	M130 ANTIGEN, CYTOPLAS	2.51e-38
7	251	32.9	1156	4	007900	M130 ANTIGEN, CYTOPLAS	2.51e-38
8	250	32.7	754	11	092175	LYSYL OXIDASE-RELATED	4.45e-38
9	243	31.8	638	4	09Y5Y8	LYSYL OXIDASE-RELATED	2.41e-36
10	243	31.8	774	4	09Y4K0	LYSYL OXIDASE-RELATED	2.41e-36
11	242	31.7	480	6	029112	SCAVENGER-RECEPTOR PRO	4.26e-36
12	242	31.7	804	6	029113	SCAVENGER-RECEPTOR PRO	4.26e-36
13	240	31.4	1785	4	09Y4V9	DMBT1/6KB.1 PROTEIN PR	1.33e-35
14	240	31.4	1785	4	09Y211	DMBT1 PROTEIN	1.33e-35
15	239	31.3	369	6	029110	SCAVENGER-RECEPTOR PRO	2.34e-35
16	239	31.3	518	11	060754	BACTERIA BINDING MACRO	4.12e-35
17	238	31.2	437	6	029109	SCAVENGER-RECEPTOR PRO	3.96e-34
18	234	30.6	600	6	028910	MUCIN (FRAGMENT)	6.97e-34
19	233	30.5	1594	6	095218	HENSLIN	3.79e-33
20	230	30.1	918	13	092098	PEMA-SRCR PROTEIN PREC	

21	225	29.5	347	4	043866	SP ALPHA.	6.29e-32
22	219	28.7	585	4	008380	MAC-2 BINDING PROTEIN	1.81e-30
23	217	28.4	483	11	09WB9	MACROPHAGE RECEPTOR MA	5.51e-30
24	217	28.4	520	4	09Y5S3	MACROPHAGE RECEPTOR	5.51e-30
25	215	28.1	2083	11	060997	CSP-DUCTIN PRECURSOR (1.68e-29
26	212	27.7	356	6	097682	UNKNOWN MRNA, PARTIAL	8.87e-29
27	212	27.7	574	11	070513	NAMA.	8.87e-29
28	212	27.7	1290	11	062827	EMERIN.	8.87e-29
29	211	27.6	822	13	09YHC1	LYSYL OXIDASE HOMOLOG	1.54e-28
30	211	27.6	895	13	09W6N1	LYSYL OXIDASE RELATED	1.54e-28
31	210	27.5	868	5	09Y1V3	TUNICATE RETINOIC ACID	2.69e-28
32	209	27.4	127	6	09Y316	M130 ANTIGEN (FRAGMENT	4.67e-28
33	206	27.0	578	6	028908	MUCIN (FRAGMENT)	2.45e-27
34	205	26.8	574	11	P70117	PANCREAS CANCER-ASSOCI	4.24e-27
35	204	26.7	665	11	061003	T-CELL DIFFERENTIATION	7.36e-27
36	202	26.4	531	5	017064	SP85.	2.21e-26
37	202	26.4	577	11	007797	PEPTIDYLPROXYL ISOMERA	2.21e-26
38	198	25.9	532	5	025111	A SPERM-ACTIVATING PEP	1.97e-25
39	194	25.4	352	11	035301	SP-ALPHA.	1.75e-24
40	194	25.4	352	11	035300	MEMBRANE PROTEIN SCAVE	8.92e-24
41	191	25.0	462	6	028881	CYCLOPHILIN C-ASSOCIAT	4.52e-23
42	191	25.0	574	11	035649	CD6C.	4.52e-23
43	188	24.6	595	4	09Y4K9	CD6D.	4.52e-23
44	188	24.6	601	4	09Y4K8	CD6B.	4.52e-23
45	188	24.6	635	4	09Y4L0		

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	097375	266	34.8	2153	5	097375	SCAVENGER RECEPTOR CYS	4.54e-42
AC	097375							
DT	01-MAY-1999 (TRENBLREL. 10, Created)							
DT	01-MAY-1999 (TRENBLREL. 10, Last sequence update)							
DT	01-NOV-1999 (TRENBLREL. 12, Last annotation update)							
DE	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.							
GN	SRCR12.							
OS	Strongylocentrotus purpuratus (Purple sea urchin).							
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;							
OC	Euechinozoa; Echinozoa; Echinoidea; Echinozoa; Echinozoa;							
OC	Strongylocentrotus.							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RA	PANCER 2., RAST J., DAVIDSON E.H.;							
RT	"Evolution of the Immune System: Transcription Factors and Downstream							
RT	Genes of the Mammalian Immune System Expressed in Sea Urchin							
RT	Coelomocytes."							
RL	Immunogenetics 0:0-0(1999).							
DR	EMBL; AF064259; AAD08654.1.							
DR	PROSITE; PS00420; SPERACT_RECEPTOR, 15.							
KW	Signal; Receptor.							
FT	SIGNAL							
FT	CHAIN							
FT	POTENTIAL.							
FT	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN							
FT	TYPE 12.							
FT	SEQUENCE							
FT	2153 AA; 226557 MW; CA3BA8A9 CRC32;							
FT	Query Match							
FT	Best Local Similarity 38.6%; Score 266; DB 5; Length 2153;							
FT	Matches 39; Conservative 22; Mismatches 40; Indels 0; Gaps 0;							
DB	462 VRLVGLNNREGVFEFLNNMGTCVDDMGTPDAVVCROLYPGSGARSAYRGSGS 521							
OY	117 VRLRGKNEFEVYASGVTGSSHMWDSASVICHQLOLGKGLAKQTPFGLGL 176							
DB	522 VPLLDVNCSCGNSLCSNNGICVHNCQHOEDASVYCT 562							
OY	177 IPIYMSVRCRDEDEMTILCERDIMOGVCYCPQMAAAVTC 217							
RESULT	2							
ID	096943							
ID	096943							
DT	01-MAY-1999 (TRENBLREL. 10, Created)							

RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
CP EMBL: Z22971; CA80544.1; -
PF PAM: PF00530; SRCR: 9; -
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1149 M130 ANTIGEN, EXTRACELLULAR VARIANT.
SQ SEQUENCE 1149 AA; 124328 MW; 4901C708 CRC32;
Query Match 32.9%; Score 251; DB 4; Length 1149;
Best Local Similarity 39.8%; Pred. No. 2,51e-38;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
957 IRLQEGTSCSGREIWHGSGMTGVCDDSDWLDLDAQVCOGLGCPALKAFAEF-GQG 1015
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
Db 1016 TGPI-WLNEVKCKGNESLWDCPARRWGSHSECGHKEADAANCT 1057
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
OY 176 LIPIYW-SNVRGRDDENILLCEKDIWOGVCPQKMAAAVTC5 217
RESULT 6
ID 007899 PRELIMINARY; PRT: 1151 AA.
AC 007899;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA LAM S.A., MICKLEK K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: Z22969; CA80542.1; -
PF PAM: PF00530; SRCR: 9; -
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1151 M130 ANTIGEN, CYTOPLASMIC VARIANT 1.
SQ SEQUENCE 1151 AA; 124820 MW; A7ZEDD2F CRC32;
Query Match 32.9%; Score 251; DB 4; Length 1151;
Best Local Similarity 39.8%; Pred. No. 2,51e-38;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
924 IRLQEGTSCSGREIWHGSGMTGVCDDSDWLDLDAQVCOGLGCPALKAFAEF-GQG 982
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
Db 983 TGPI-WLNEVKCKGNESLWDCPARRWGSHSECGHKEADAANCT 1024
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
OY 176 LIPIYW-SNVRGRDDENILLCEKDIWOGVCPQKMAAAVTC5 217
RESULT 7
ID 007900 PRELIMINARY; PRT: 1156 AA.
AC 007900;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 2 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA LAM S.A., MICKLEK K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
CP EMBL: Z22970; CA80543.1; -
PF PAM: PF00530; SRCR: 9; -
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1156 M130 ANTIGEN, CYTOPLASMIC VARIANT 2.
SQ SEQUENCE 1156 AA; 125352 MW; 287A07A0 CRC32;
Query Match 32.9%; Score 251; DB 4; Length 1156;
Best Local Similarity 39.8%; Pred. No. 2,51e-38;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
924 IRLQEGTSCSGREIWHGSGMTGVCDDSDWLDLDAQVCOGLGCPALKAFAEF-GQG 982
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
OY 117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
Db 983 TGPI-WLNEVKCKGNESLWDCPARRWGSHSECGHKEADAANCT 1024
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
OY 176 LIPIYW-SNVRGRDDENILLCEKDIWOGVCPQKMAAAVTC5 217
RESULT 8
ID 092175 PRELIMINARY; PRT: 754 AA.
AC 092175;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, last sequence update)
DT 01-AUG-1999 (TEMBLrel. 11, last annotation update)
DE LYSYL OXIDASE-RELATED PROTEIN 2 (LOR2 PROTEIN).
GN LOR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J AND 129/SVJ; TISSUE=MUSCLE;
MEDLINE: 99126643.
RA JANG W., HUN A., SPILSON S.V., MILLER W., ROE B.A., MEISLER M.H.;
RT "Comparative sequence of human and mouse BAC clones from the mnd2
region of chromosome 2p13.";
RL Genome Res. 9:53-61(1999).
DR EMBL: AF053368; AAC83205.1; -
DR EMBL: AF084363; AAC95338.1; -
SQ SEQUENCE 754 AA; 83681 MW; 668B44B2 CRC32;
Query Match 32.7%; Score 250; DB 11; Length 754;
Best Local Similarity 40.2%; Pred. No. 4,45e-38;
Matches 41; Conservative 20; Mismatches 39; Indels 2; Gaps 2;
308 VRLKGAGHGEGRVEYVAKATGWTCVDRKWDLQAASYVCELFQFARRELSGARNGQM 367
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHLQLGKGLA-KQTFPSGLG 175
Db 368 GAHLSEVRCSGOEPLMRCPKNI-TAEDCSHSDAGVGN 408
117 IPIYWSNVRGRDDENILLCEKDIWOGVCPQKMAAAVTC5 217
OY 176 LIPIYW-SNVRGRDDENILLCEKDIWOGVCPQKMAAAVTC5 217
RESULT 9
ID 09518 PRELIMINARY; PRT: 638 AA.
AC 09518;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, last annotation update)
DE LYSYL OXIDASE-LIKE PROTEIN 2.


```
RESULT 13
ID 09Y4V9 PRELIMINARY: PRT: 1785 AA.
AC 09Y4V9;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE DMBT1/6KB.1 PROTEIN PRECURSOR.
GN DMBT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE: 97434209.
MOLENAUER J., WIEMANN S., SCHEURLEN W., KORN B., HAYASHI Y.,
WILGENBUS K.K., VON DEIMLING A., POUSTKA A.;
"DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-
26.1 is deleted in malignant brain tumours.";
Nat. Genet. 17:32-39(1997).
DR EMBL: A0000342; CAA04019.1;
RL SIGNAL.
FT CHAIN 1 25 POTENTIAL.
SQ SEQUENCE 1785 AA; 193941 MW; 9259828E CRC32;

Query Match 31.4%; Score 240; DB 4; Length 1785;
Best Local Similarity 38.2%; Pred. No. 1.33e-35;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

Db 623 LRLVNGDRCQGVVEVLYGSGWGTCDYDNDANVYCRQLCGGAMSPGNARF-GQG 661
QY 117 VRLRGKNEEGVEVYAGWGTVCSSHWDDSDASVICHQQLG-GKGIKQTPSSGLG 175
DB 682 SCPIVDVRCSGHESYLWSCPHNGWLSHNGCHHEDAGVICS 723
QY 176 LPIYNSVNRCDENILLCERDINQGVCPQKMAAYTCS 217

RESULT 14
ID 09Y211 PRELIMINARY: PRT: 1785 AA.
AC 09Y211;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE DMBT1 PROTEIN.
GN DMBT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC HORTI A.;
RX DMBT1.;
RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB020812; BAA78577.1;
DR EMBL: AB020812; BAA78577.1; JOINED.
DR EMBL: AB020813; BAA78577.1; JOINED.
DR EMBL: AB020813; BAA78577.1; JOINED.
DR EMBL: AB020814; BAA78577.1; JOINED.
DR EMBL: AB020815; BAA78577.1; JOINED.
DR EMBL: AB020816; BAA78577.1; JOINED.
DR EMBL: AB020817; BAA78577.1; JOINED.
DR EMBL: AB020818; BAA78577.1; JOINED.
DR EMBL: AB020819; BAA78577.1; JOINED.
DR EMBL: AB020820; BAA78577.1; JOINED.
DR EMBL: AB020821; BAA78577.1; JOINED.
DR EMBL: AB020822; BAA78577.1; JOINED.
DR EMBL: AB020823; BAA78577.1; JOINED.
DR EMBL: AB020824; BAA78577.1; JOINED.
DR EMBL: AB020825; BAA78577.1; JOINED.
DR EMBL: AB020826; BAA78577.1; JOINED.
DR EMBL: AB020827; BAA78577.1; JOINED.
DR EMBL: AB020828; BAA78577.1; JOINED.
```

```
DR EMBL: AB020829; BAA78577.1; JOINED.
DR EMBL: AB020830; BAA78577.1; JOINED.
DR EMBL: AB020831; BAA78577.1; JOINED.
DR EMBL: AB020832; BAA78577.1; JOINED.
DR EMBL: AB020833; BAA78577.1; JOINED.
DR EMBL: AB020834; BAA78577.1; JOINED.
DR EMBL: AB020835; BAA78577.1; JOINED.
DR EMBL: AB020836; BAA78577.1; JOINED.
DR EMBL: AB020837; BAA78577.1; JOINED.
DR EMBL: AB020838; BAA78577.1; JOINED.
DR EMBL: AB020839; BAA78577.1; JOINED.
DR EMBL: AB020840; BAA78577.1; JOINED.
DR EMBL: AB020841; BAA78577.1; JOINED.
DR EMBL: AB020842; BAA78577.1; JOINED.
DR EMBL: AB020843; BAA78577.1; JOINED.
DR EMBL: AB020844; BAA78577.1; JOINED.
DR EMBL: AB020845; BAA78577.1; JOINED.
DR EMBL: AB020846; BAA78577.1; JOINED.
DR EMBL: AB020847; BAA78577.1; JOINED.
DR EMBL: AB020848; BAA78577.1; JOINED.
DR EMBL: AB020849; BAA78577.1; JOINED.
DR EMBL: AB020850; BAA78577.1; JOINED.
SQ SEQUENCE 1785 AA; 193991 MW; 479F75D8 CRC32;

Query Match 31.4%; Score 240; DB 4; Length 1785;
Best Local Similarity 38.2%; Pred. No. 1.33e-35;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

Db 623 LRLVNGDRCQGVVEVLYGSGWGTCDYDNDANVYCRQLCGGAMSPGNARF-GQG 661
QY 117 VRLRGKNEEGVEVYAGWGTVCSSHWDDSDASVICHQQLG-GKGIKQTPSSGLG 175
DB 682 SCPIVDVRCSGHESYLWSCPHNGWLSHNGCHHEDAGVICS 723
QY 176 LPIYNSVNRCDENILLCERDINQGVCPQKMAAYTCS 217

RESULT 15
ID 029110 PRELIMINARY: PRT: 369 AA.
AC 029110;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE SCAVENGER-RECEPTOR PROTEIN (FRAGMENT).
GN WCL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SLAB/B;
RX MEDLINE: 97364683.
KA KANAN J.H.C., NAYEM N., BINNS R.M., CHAIN B.M.;
"Mechanisms for variability in a member of the scavenger-receptor
RT cytochrome-rich superfamily";
ImmunoGenetics 46:276-282(1997).
RL EMBL: X99333; CAA67707.1;
DR PROSITE: PS00420; SPBRCT_RECEPTOR. 2.
DR PFM: PF00530; SRCR; 3.
FT NON_TER 1 369
FT NON_TER 369 369
SQ SEQUENCE 369 AA; 39710 MW; C669BB4B CRC32;

Query Match 31.3%; Score 239; DB 6; Length 369;
Best Local Similarity 37.9%; Pred. No. 2.34e-35;
Matches 39; Conservative 20; Mismatches 40; Indels 4; Gaps 4;

Db 259 LRLRGDSVCSGVVEVYHSGWGTCDSDSLAEVYVCOQLGCHALPDLDAF-GPG 317
QY 117 VRLRGKNEEGVEVYAGWGTVCSSHWDDSDASVICHQQLG-GKGIKQTPSSGLG 175
DB 318 SGI-WLDEVQCRGRSSLMDCAGPWSGSDCKHEDAGVICS 359
```

Thu Mar 16 07:57:51 2000

US-09-147-947-6-03.rspt

Page 6

OY 176 LPIYW-SNVRCRDEENILCEKDIWQSGVCPQKMAAVTCS 217

Search completed: Mon Mar 13 10:24:48 2000
Job time : 19 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

ch-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:28:46 2000; Maspar time 6.74 Seconds
Tabular output not generated. 355.087 Million cell updates/sec

Title: >US-09-147-947-6
Description: (227-327) from US09147947A-pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLAGSSVHEGRVELYHAG.....KSMGHRNGHKEADGVSC 101

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1-geneseqp

Statistics: Mean 27.982; Variance 104.818; scale 0.267

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

lt	Score	Query Match	Length	ID	Description	Pred. No.
1	757	100.0	822	1 W90987	Human serine protease	1.56e-71
2	757	100.0	875	1 W83361	Human neutrolypsin.	4.88e-66
3	708	93.5	761	1 W83362	Mouse neutrolypsin.	4.88e-66
4	708	93.5	761	1 W90988	Mouse serine protease	4.88e-66
5	435	57.5	453	1 R27036	Bovine sol. scavenger	9.57e-36
6	435	57.5	453	1 R05509	Scavenger receptor pro	9.57e-36
7	428	56.5	451	1 R40802	Human scavenger recept	5.59e-35
8	428	56.5	451	1 R27035	Human sol. scavenger r	5.59e-35
9	428	56.5	451	1 W08077	Type I macrophage scav	5.59e-35
10	427	56.4	451	1 R21512	Human macrophage recept	7.19e-35
11	427	56.4	451	1 W19708	Macrophage scavenger r	7.19e-35
12	412	54.4	585	1 R40172	Sequence of an immunor	3.13e-33
13	412	54.4	585	1 R40213	Sequence of a 90k tumo	3.13e-33
14	412	54.4	585	1 R36533	Gp85-97 clone 18 prod	3.13e-33
15	412	54.4	585	1 W81363	Human cytokine regulat	3.13e-33
16	412	54.4	585	1 R41359	Tumour associated 90K	3.13e-33
17	402	53.1	666	1 W64590	Human SRCR protein fra	3.87e-32
18	402	53.1	1785	1 W64591	Human SRCR protein fra	3.87e-32
19	373	49.3	574	1 R40784	Sequence of cyclophilli	5.51e-29
20	373	49.3	574	1 W01632	Mouse cytokine regulat	5.51e-29
21	365	48.2	1290	1 W07609	Rat von Ebner's gland	4.06e-28
22	356	47.0	347	1 W68200	Human scavenger recept	3.82e-27
23	356	47.0	347	1 Y13369	Amino acid sequence of	3.82e-27

24	356	47.0	347	1 W64537	Human liver cell clone	3.82e-27
25	355	46.9	489	1 W39749	Mouse macroSR protein.	4.91e-27
26	355	46.9	518	1 W03561	Macrophage receptor wi	4.91e-27
27	338	44.6	495	1 W39747	Human macroSR protein	3.35e-25
28	338	44.6	520	1 W39748	Human macroSR protein	3.35e-25
29	338	44.6	520	1 Y04374	Human macrophage recep	3.35e-25
30	125	16.5	798	1 R57283	Bovine enterokinase.	3.04e-03
31	112	14.8	349	1 W35849	Human CD5 for use in F	4.93e-02
32	88	11.6	356	1 W46917	Amino acid sequence of	6.74e+00
33	83	11.0	652	1 W08010	S. pneumoniae liase.	1.79e-01
34	83	11.0	1167	1 W17700	CryET4.	1.79e-01
35	83	11.0	1167	1 W35258	Bacillus thuringiensis	1.79e-01
36	83	11.0	1167	1 W87632	CryET4 protein sequenc	1.79e-01
37	83	11.0	1167	1 R54073	CryET4.	1.79e-01
38	81	10.7	323	1 R54944	Cuphea hookeriana thlo	2.63e+01
39	80	10.6	1722	1 W38429	Human dendritic cell r	3.18e-01
40	80	10.6	1723	1 W00645	Mouse DEC-205.	3.18e-01
41	80	10.6	1743	1 W00644	Human DEC-205.	3.18e-01
42	79	10.4	1168	1 R89493	CryIF class toxin 91C2	3.85e+01
43	77	10.2	980	1 W69741	SAPAP2 protein.	5.62e+01
44	77	10.2	980	1 W69743	SAPAP1 protein.	5.62e+01
45	77	10.2	1130	1 R13436	Merosin M polypeptide.	5.62e+01

ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	W90987	standard; Protein; 822 AA.
AC	W90987	
DT	13-MAY-1999	(first entry)
DE	Human serine protease BSSP-3.	
KW	Serine protease; BSSP-3; brain tissue.	
OS	Homo sapiens.	
PN	W09905290-A1.	
PD	04-FEB-1999.	
PE	24-JUL-1998; J03324.	
PR	24-JUL-1997; JP-213969.	
PA	(SUNR) SUNTOXY LTD.	
PI	Tsuruoka N, Yamaguchi N, Yamashiro K;	
DR	WPI; 99-142942/12.	
DR	N-PSDE; X19024.	
PT	New serine protease expressed in brain tissue - used in screening	
PF	for potential serine protease inhibitors for drug use	
PS	Claim 1: Page 61-65; 69pp; Japanese.	
CC	The present sequence is a serine protease designated BSSP-3, which	
CC	is isolated from human brain tissue. Transformants may be used to	
CC	produce the enzyme or its partial sequences. Products from the present	
CC	invention are used for screening for potential peptide or non-peptide	
CC	serine protease inhibitors or expression regulators for use as drugs.	
SO	Sequence 822 AA;	
Query Match	100.0%; Score 757; DB 1; Length 822;	
Best Local Similarity	100.0%; Pred. No. 1.56e-71;	
Matches	101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	227 IRLAGSSVHEGRVELYHAGOMGTVCDDMDADAEVTCRLGSLGAKAHQAFEGS 286	
Qy	227 IRLAGSSVHEGRVELYHAGOMGTVCDDMDADAEVTCRLGSLGAKAHQAFEGS 286	
Db	287 GPVMLDEVRCGTGNELSTEQCRKSSWGSHNGHKEADGVSC 327	
Qy	287 GPVMLDEVRCGTGNELSTEQCRKSSWGSHNGHKEADGVSC 327	
RESULT	2	
ID	W83361	standard; Protein; 875 AA.
AC	W83361	
DT	17-FEB-1999	(first entry)
DE	Human neutrolypsin.	
KW	Human; neutrolypsin; tumour inhibition; neurological disease;	
KW	lung disease; gene therapy; drug development; stroke; brain injury;	
KW	neurodegeneration; neuroinflammatory disease; multiple sclerosis;	
KW	epilepsy; hypoxia; ischaemia; nerve transection; neovangiogenesis;	

KW emphysema; bronchitis.
 OS Homo sapiens.
 PN MO9849322-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998: IB0625.
 PR 26-APR-1997: CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sondegger P.
 DR MPI: 99-009438/01.
 DR N-PSDB: V72589.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 20-24: 50pp: English.
 CC The present sequence represents human neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 875 AA:

Query Match 100.0%; Score 757; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 1,56e-71;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 280 IRLAGSSVHGRVETLHAGWGTCDDQMDADAEVTCRQLGSLGAKAMHQAFFEGS 339
 QY 227 IRLAGSSVHGRVETLHAGWGTCDDQMDADAEVTCRQLGSLGAKAMHQAFFEGS 286
 Db 340 GPVNLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 380
 QY 287 GPVNLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 327

RESULT 3
 ID W83362 standard; Protein: 761 AA.
 AC W83362;
 DT 17-FEB-1999 (first entry)
 PT Mouse neurotysin.

Mouse: neurotysin: tumour inhibition; neurological disease;
 lung disease; gene therapy; drug development; stroke; brain injury;
 neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 epilepsy; hypoxia; ischaemia; nerve transection; neovascularisation;
 emphysema; bronchitis.

OS Mus musculus.
 PN MO9849322-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998: IB0625.
 PR 26-APR-1997: CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sondegger P.
 DR MPI: 99-009438/01.
 DR N-PSDB: V72590.

PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 29-32: 50pp: English.
 CC The present sequence represents mouse neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 761 AA:

Query Match 93.5%; Score 708; DB 1; Length 761;
 Best Local Similarity 91.0%; Pred. No. 4,88e-66;
 Matches 91; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 166 IRLVGNNGHGRVETLHAGWGTCDDQMDADAEVTCRQLGSLGAKAMHQAFFEGS 225
 QY 227 IRLAGSSVHGRVETLHAGWGTCDDQMDADAEVTCRQLGSLGAKAMHQAFFEGS 286
 Db 226 GPILLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 265
 QY 287 GPVNLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 326

RESULT 4
 ID W99088 standard; Protein: 761 AA.
 AC W99088;
 DT 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 KW Serine protease; BSSP-3; brain tissue.
 OS Mus sp.
 PN WC9905290-A1.
 PD 04-FEB-1999.
 PF 24-JUL-1998: J03324.
 PR 24-JUL-1997: JP-215969.
 PA (SONR) SUNTORY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR MPI: 99-142942/12.
 DR N-PSDB: X19027.

PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1: Page 51-54: 69pp: Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transformants may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 761 AA:

Query Match 93.5%; Score 708; DB 1; Length 761;
 Best Local Similarity 91.0%; Pred. No. 4,88e-66;
 Matches 91; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 166 IRLVGNNGHGRVETLHAGWGTCDDQMDADAEVTCRQLGSLGAKAMHQAFFEGS 225
 QY 227 IRLAGSSVHGRVETLHAGWGTCDDQMDADAEVTCRQLGSLGAKAMHQAFFEGS 286
 Db 226 GPILLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 265
 QY 287 GPVNLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 326

RESULT 5

ID R27036 standard; Protein: 453 AA.
 AC R27036;
 DT 16-FEB-1993 (first entry)
 DE Bovine sol. scavenger receptor.
 KW Macrophage; endotoxaemia; radiolabelled; toxic; degenerate.
 OS Bos taurus.
 FH Key
 FT domain 1. 50

FT	domain	/note= "cytoplasmic"	51.76
FT	domain	/note= "transmembrane"	77.109
FT	domain	/note= "spacer"	110.271
FT	domain	/note= "alpha helical coiled coil"	272.343
FT	domain	/note= "collagen binding"	341.451
FT	domain	/note= "Cys-rich"	
PN	<u>MO214482-A</u>		
PD	03-SEP-1992.		
PE	21-FEB-1992; U01370.		
PR	22-FEB-1991; US-662227.		
	(MASI) MASSACHUSETTS INST TECHNOLOGY.		
	Krieger M;		
	WPI: 92-315935/38.		
	N-PSDB: Q28540.		
PT	Treatment and diagnosis of endotoxaemia and related disease		
PT	states - using a polypeptide fragment of the extracellular		
PT	portion of a macrophage scavenger receptor protein		
PS	Disclosure: Page 38; 56pp; English.		
CC	The bovine scavenger receptor protein sequence was deduced from the		
CC	DNA sequence obtd. by screening a bovine lung cDNA library with		
CC	degenerate probes designed based on fragments of the purified		
CC	scavenger receptor protein. The scavenger receptor protein has a		
CC	binding capacity for acetylated low density lipoprotein (LDL) and		
CC	binds endotoxin. Fragments of the extracellular portion of the		
CC	scavenger receptor protein inactivate endotoxin-related substances		
CC	and are used in the diagnosis and treatment of endotoxaemia. They		
CC	may be administered to patients at high risk of symptomatic or		
CC	endotoxic shock. The fragments may also be fixed to inert supports		
CC	for purification purposes. The fragments may be modified to have		
CC	greater binding affinity for the endotoxin-related substance than		
CC	the native scavenger receptor protein, or to more effectively		
CC	neutralise the toxic or pathogenic effects of mols. that bind		
CC	the scavenger receptor protein or of organisms which express such mols.		
CC	See also R27035.		
CC	Sequence 453 AA;		
QY	Query Match	57.5%; Score 435; DB 1; Length 453;	
	Best Local Similarity 54.5%; Pred. No. 9.57e-36;		
	Matches 55; Conservative 18; Mismatches 28; Indels 0; Gaps 0		
Db	352 VRLVGGSPHGHRVDFIEFGOMGTCDDDRWEIRGLVLCRSITGKYGVSVKRAYFGKGT 411		
	::: : ::: ::: :: ::: :		
	227 IRLAGGVSVHGSRVLYHAGOMGTCDQDQMDADAEVIEIRQLGSLAKAHQAATFGSGS 286		
	412 GPIWINEVFCFPKSSIEECRIKRWGVACSHDEDAVYTC 452		
	::: ::: ::: :: :: ::		
QY	287 GPVMDIEVRCGTGNELSIQCPSKSGEHNCGHKEDAGVSC 327		
RESULT	6		
ID	R05509 standard; protein; 453 AA.		
AC	R05509.		
DT	23-OCT-1990 (first entry)		
DE	Scavenger receptor protein with affinity for acetylated low density		
DE	lipoprotein (aLDL).		
KW	Acylated low density lipoprotein; aLDL; atherosclerotic plaque;		
KW	ds.		
PN	WO9005748-A.		
PD	31-MAY-1990.		
PE	14-NOV-1989; 005116.		
PR	15-NOV-1988; US-272002.		
PA	(MASI) Massachusetts Inst Tech.		
P1	Engelkner TJ;		
DR	WPI: 90-193408/25.		
DR	N-PSDB: 004926.		
PT	New receptor protein -		
PT	has affinity for acetylated low density lipoprotein and		
PT	corresponding antibodies and DNA sequences.		

PS Disclosure: 79pp; English.

CC Receptor protein, and fragments and analogues thereof may be immobilised on a support and used in assay and purification of the ADL target. Labelled Abs, raised to the protein may be injected into the vascular system to detect the presence of atherosclerotic plaques.

CC Sequence 453 AA;

S0

Query Match 57.5%; Score 435; DB 1; Length 453;
Best Local Similarity 54.5%; Pred. No. 9.57e-36;
Matches 55; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

Db 352 VRLVGGSPHFGREIIFHEGOMGVCDDBRELRCGLIVCSLSLKGYQSHKRAHFKECT 411
Y 227 IRLVGSSVHGRELXELHAGQMGVCDWDADAELICQLSLGIKAWHOAYFEEGS 286
D 412 GPIWLNEVFCEPKRESSIEECRIROGWGRACSHDEDAVGTCT 452
Y 287 GPVWLDVEYRGTGNELSTIEQCCKSSWGEBHGNCHEKDAGVSCCT 327

RESULT 7
ID R40802 standard; Protein; 451 AA.
AC R40802:
DT 23-FEB-1994 (first entry)
DE Human scavenger receptor I.
KW Human; scavenger receptor; antihuman scavenger receptor antibody;
KW AHRSA; mammal; rabbit; antiserum; monoclonal; HAT-resistant hybridoma;
OS arteriosclerosis.
OS Homo sapiens.
PN J0519217-A.
PD 03-AUG-1993.
PF 16-AUG-1991; 229728.
PR 27-AUG-1990; JP-223398.
PA (CHUS.) CHUGAI PHARM CO LTD.
DR WPI: 93-277488/35.
DR N-PDSB: Q47730.
PT Anti-human scavenger receptor antibody - useful for monitoring progress of arteriosclerosis
PT Antigen of arteriosclerosis
PS Claim 2; Page 5-7; 10pp; Japanese.
CC The sequences given in R40802-03 represent the human scavenger receptors I and II. These proteins may be used in the production of antihuman scavenger receptor antibodies (AHSRA). These peptides were used to immunise a mammal, pref. a rabbit, and anti-serum was prepared. A monoclonal antibody was prepared from HAT-resistant hybridoma. These antibodies may be used to determine the progress of arteriosclerosis.

S0 Sequence 451 AA;

Query Match 56.5%; Score 428; DB 1; Length 451;
Best Local Similarity 53.5%; Pred. No. 5.59e-35;
Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSPHFGREIIFHEGOMGITCDDREVRVYGVCNLSLEPGVQAHHKAHNGCGT 409
Y 227 IRLVGSSVHGRELXELHAGQMGVCDWDADAELICQLSLGIKAWHOAYFEEGS 286
D 410 GPIWLNEVFCEPKRESSIEECKIRQMGTRACSHSEDAVGTCT 450
Y 287 GPVWLDVEYRGTGNELSTIEQCCKSSWGEBHGNCHEKDAGVSCCT 327

RESULT 8
ID R27035 standard; Protein; 451 AA.
AC R27035:
DT 16-FEB-1993 (first entry)
DE Human scavenger receptor.
KW Macrophage; endothaxemia; radiolabelled; toxic; degenerate.
OS Homo sapiens.
FH Key location/Qualifiers
FT domain 1..50
FT /note= "cytoplasmic"
FT domain 51..76

FT domain /note="transmembrane"
FT 77..109
FT /note="spacer"
FT 110..271
FT domain /note="alpha helical coiled coil"
FT 272..343
FT /note="collagen binding"
FT 341..451
FT /note="Cys-rich"
PN W09214482-A.
PD 03-SEP-1992.
PF 21-FEB-1992; U01370.
PR 22-FEB-1991; US-662227.
PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
PI Krieger M;
DR WPI; 92-315935/38.
N-PSDB: Q28539.
Treatment and diagnosis of endotoxaemia and related disease states - using a polypeptide fragment of the extracellular portion of a macrophage scavenger receptor protein
Disclosure; Page 33; 56pp; English.
The human scavenger receptor protein sequence was deduced from the DNA sequence obt'd. by screening a platelet cDNA library with degenerate probes designed based on fragments of the purified scavenger receptor protein. The scavenger receptor protein has a binding capacity for acetylated low density lipoprotein (LDL) and binds endotoxin. Fragments of the extracellular portion of the scavenger receptor protein inactivate endotoxin-related substances and are used in the diagnosis and treatment of endotoxaemia. They may be administered to patients at high risk of symptomatic or endotoxic shock. The fragments may also be fixed to inert supports for purification purposes. The fragments may be modified to have greater binding affinity for the endotoxin-related substance than the native scavenger receptor protein, or to more effectively neutralise the toxic or pathogenic effects of mols. that bind to the scavenger receptor protein or of organisms which express such mols.
See also R27036.
Sequence 451 AA:
SQ

Query Match 56.5%; Score 428; DB 1; Length 451;
Best Local Similarity 53.5%; Pred. No. 5.59e-35;
Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSGHEGRVVELHSGQWGTICDDRMVEYRGVQVCRSLGYGVQAVHKAHFGGCT 409
QY 227 IRLAGSSVHEGRVVELYHAGQWGTICDDMDADAEVLCRQLGSLGAKAHQAVFGECS 286

Db 410 GPIWLNEVFCFGRESSIECKIRQWGTFRACSHSDAGVTCT 450
QY 287 GPVMDDEVCTGNELSTIEQCPKSSWGEHNGCHKEDAGVSCCT 327

AC W08077 standard; Protein: 451 AA.
ID W08077.
AC W08077.

DE 25-FEB-1997 (first entry)
DE Type I macrophage scavenger receptor.
KW Type I macrophage scavenger receptor; SRG; kidney; 293 cell;
KW cell substrate; cell attachment; atherosclerosis.

OS Homo sapiens.
PN W09638725-A1.
PD 05-DEC-1996.

PF 30-MAY-1996; U08081.
PR 30-MAY-1995; US-453117.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Brawner ME; Elshoutbagy NAE, Lysko PG;
DR WPI; 97-034505/03.
N-PSDB: T43259.

PT New modified human embryonic kidney 293 cells - transfected with a mammalian scavenger receptor gene to enhance ability to attach to a solid support
PT Disclosure: Page 10-13; 32pp; English.
PS Human type I (W08077) and type II (W08078) macrophage scavenger

CC receptors can be expressed in human embryonic kidney 293 cells
CC following transfection of the cells with vectors carrying the
CC respective scavenger receptor genes (T43259 and T43260). The
CC transfected cells show an enhanced ability to attach to a solid
CC support and are useful for screening cpds. for biological activity,
CC or for identifying antagonists of the scavenger receptor gene, e.g.
CC to develop agents for treatment of atherosclerosis.
SQ Sequence 451 AA;

Query Match 56.5%; Score 428; DB 1; Length 451;
Best Local Similarity 53.5%; Pred. No. 5.59e-35;
Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSGHEGRVVELHSGQWGTICDDRMVEYRGVQVCRSLGYGVQAVHKAHFGGCT 409
QY 227 IRLAGSSVHEGRVVELYHAGQWGTICDDMDADAEVLCRQLGSLGAKAHQAVFGECS 286

Db 410 GPIWLNEVFCFGRESSIECKIRQWGTFRACSHSDAGVTCT 450
QY 287 GPVMDDEVCTGNELSTIEQCPKSSWGEHNGCHKEDAGVSCCT 327

DE 19-MAY-1992 (first entry)
DE Human scavenger receptor type I.
KW Lipoproteins.
OS Homo sapiens.
PN J03290184-A.
PD 19-DEC-1991.

PF 06-APR-1990; 090274.
PR 06-APR-1990; JP-090274.
PA (CHUS) CHUGAI PHARMACEUTICAL KK.
DR WPI; 92-051436/07.
N-PSDB: Q21549.

PT New scavenger receptor-producing animal cells - which have been
PT transformed with vector contg. gene for coding human scavenger
PT receptor I or II type under control of promoter, etc.
PT Disclosure; Fig 1; 9pp; Japanese.

CC The gene encoding the receptor can be expressed from a cytomegalo-
CC virus promoter in a host cell e.g. CHO cell. The resulting
CC recombinant scavenger receptor can be used to detect modified lipo-
CC proteins or modified substances in the blood.
CC See also R21513.

CC Sequence 451 AA;
SQ

Query Match 56.4%; Score 427; DB 1; Length 451;
Best Local Similarity 52.5%; Pred. No. 7.19e-35;
Matches 53; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSGHEGRVVELHSGQWGTICDDRMVEYRGVQVCRSLGYGVQAVHKAHFGGCT 409
QY 227 IRLAGSSVHEGRVVELYHAGQWGTICDDMDADAEVLCRQLGSLGAKAHQAVFGECS 286

Db 410 GPIWLNEVFCFGRESSIECKIRQWGTFRACSHSDAGVTCT 450
QY 287 GPVMDDEVCTGNELSTIEQCPKSSWGEHNGCHKEDAGVSCCT 327

AC W19708 standard; Protein: 451 AA.
ID W19708.
AC W19708.

DE 19-AUG-1997 (first entry)
DE Macrophage scavenger receptor protein.
KW LTA; Gram-positive bacteria; cell wall; atherosclerosis; host defence;
KW septicemia; inhibitor; complement activation; cytokine release; therapy;
KW nitric oxide production; bacterial infection; septic shock.

OS Homo sapiens.
PN US5624904-A.
PD 29-APR-1997.

PT 17-NOV-1993; 154365.
PT Disclosure: Page 10-13; 32pp; English.
PS Human type I (W08077) and type II (W08078) macrophage scavenger

PR 17-NOV-1993: US-154365.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI (UYA) UNIV YALE.
 PI Joiner KA, Krieger M;
 DR WPI; 97-258236/23.
 N-PSDB: T68796.
 PT Treatment of septicemia caused by Gram-positive bacteria - by
 administration of macrophage scavenger receptor protein
 PS Disclosure; Column 23-26; 16pp; English.
 CC This sequence represents the human macrophage scavenger receptor protein
 (MSRP). MSRP binds specifically to the lipoteichoic acid (LTA) residues
 on the gram-positive bacterial cell wall. MSRP exhibit unusually broad
 binding specificity for polyanionic ligands, and have been implicated in
 atherosclerosis and a variety of host defence functions. This protein can
 be used in the method of the invention. The method of the invention is
 for the treatment of septicemia caused by Gram-positive bacteria. The
 method comprises administration of a MSRP in an amount sufficient to
 inhibit complement activation, cytokine release or nitric oxide
 production induced by LTA released by the Gram-positive bacteria. In
 addition to treating Gram-positive septic shock, the MSRP can be used to
 screen for other compounds for treating Gram-positive septic shock. MSRP
 can also be used to purify, label or detect LTA or LTA-containing cells.
 CC MSRP, active MSRP fragments, anti-MSRP antibodies or other compounds that
 inhibit binding of Gram-positive bacteria to MSRP can be used to treat
 CC pathologies such as septicemia, Gram-positive bacterial infection,
 CC Gram-positive septicemia or Gram-positive or Gram-negative septic shock.
 SQ Sequence 451 AA;

Query Match 56.4%; Score 427; DB 1; Length 451;
 Best Local Similarity 53.5%; Pred. No. 7,136-35; Indels 0; Gaps 0;
 Matches 54; Conservative 19; Mismatches 28;

Db 350 VRLVGGSGPREGREYELHSGQWGTICDNNVEYGVQVCRSLGPGVQAAHKAHFGCGT 409
 :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 227 IRLAGSSVHEGRVELYHAGQWGTICDDQMDADAETICQLGSLGAKMHWQAYFEEGS 266
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 410 GPWLMEVQCTGEASLADCKSLGWLKSNCRHBRDAGVCT 450
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 287 GPWLMEVQCTGEASLADCKSLGWLKSNCRHBRDAGVCT 327

RESULT 12
 ID R40172 standard; Protein; 585 AA.
 AC R40172;
 DT 08-FEB-1994 (first entry)
 DE Sequence of an immunoregulin (IR)-95 polypeptide.
 PI Tumour associated antigen; breast cancer; cell line CG-5;
 PI ovarian cancer; immunoregulin-95; IR-95.
 Key Homo sapiens.
 FT peptide Location/Qualifiers
 FT 1..17 /label= signal
 FT region 24..125
 FT /label= SCRC homology region
 FT modified_site 71
 FT /label= glycosylation site
 FT /note= "see also AAs 125,192,362,398,551,580"
 PN MO9316180-A.
 PD 19-AUG-1993.
 PF 17-FEB-1993; E00382.
 PR 17-FEB-1992; IT-RM0100.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (UYCH-) UNIV CHIETI ANNUNIO G D.
 PI Azam M, Iacobelli S, Natoli C, Sures I, Ullrich A;
 DR WPI; 93-272884/34.
 N-PSDB: Q46888.
 PT Recombinant DNA - encoding tumour associated antigen,
 PT immunoregulin-95
 PS Disclosure; Fig 1; 69pp; English.
 CC Immunoregulin-95 is a 90k tumour-associated antigen purified from
 CC the culture fluid of the human breast cancer cell line, CG-5, the
 CC serum of a breast cancer patient, or the ascitic fluid from an
 CC ovarian cancer patient. The native antigen, which has a mol. wt. of

CC 95 kD, is present as a high mol. wt. complex. Homology in the region
 CC of AAs 35-80 of the 90k antigen is found with type I macrophage
 CC scavenger receptor; sea urchin speract receptor; and human
 CC lymphocyte glycoprotein 11/Leu-1.
 SQ Sequence 585 AA;

Query Match 54.4%; Score 412; DB 1; Length 585;
 Best Local Similarity 53.5%; Pred. No. 3,136-33; Indels 0; Gaps 0;
 Matches 54; Conservative 19; Mismatches 28;

Db 24 MRLADGATNOGRVEYRGQWGTICDNNVEYGVQVCRSLGPGVQAAHKAHFGCGS 83
 :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 227 IRLAGSSVHEGRVELYHAGQWGTICDDQMDADAETICQLGSLGAKMHWQAYFEEGS 266
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 84 GPWLMEVQCTGEASLADCKSLGWLKSNCRHBRDAGVCT 124
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 287 GPWLMEVQCTGEASLADCKSLGWLKSNCRHBRDAGVCT 327

RESULT 13
 ID R40213 standard; Protein; 585 AA.
 AC R40213;
 DT 03-FEB-1994 (first entry)
 DE Sequence of a 90k tumour-associated antigen, IR-95, capable
 DE of binding to SP-2 (Accession No. I-1083).
 KW Tumour-associated antigen; breast cancer cell line CG-5;
 KW 90 K antigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 1..18 /label= signal
 FT region 24..125
 FT /label= SCRC homology region
 FT binding_site 69
 FT /label= potential glycosylation site
 FT /note= "see also AAs 125,192,363,398,551,580"
 PN MO9316181-A.
 PD 19-AUG-1993.
 PF 17-FEB-1993; E00385.
 PR 17-FEB-1992; IT-RM0099.
 PA (UYCH-) UNIV CHIETI ANNUNIO G D.
 PI (UYNY) UNIV NEW YORK STATE.
 PI Iacobelli S, Natoli C, Schlessinger J;
 DR WPI; 93-272885/34.
 N-PSDB: Q48170.
 PT New 90k tumour-associated antigen, IR-95 - is for use in
 PT diagnosis and therapy of cancer, HIV and auto-immune diseases
 PS Claim 2; pages 45-48; 68pp; English.
 CC 90K antigen has an apparent molecular weight of approx. 95 kd. It is
 CC a tumour associated antigen and is elevated in the serum of patients
 CC with cancer and also in patients with HIV. It reacts with Mab SP-2
 CC which was prod. by immunising mice with proteins that had been
 CC released into tissue culture fluid by human MCF-7 breast cancer
 CC cells. Mab SP-2 cell line is deposited at the Institut Pasteur,
 CC Paris, Accession number 1-1083. 90K is also present in normal
 CC subjects. It is purified from the culture fluid of the human breast
 CC cancer cell line, CG-5. N-terminal sequencing of the 90K antigen
 CC was used to design a 'guesser' nucleotide sequence (Q48171) as a
 CC probe to screen a lambda-gt10 library prepd. from MCF7 polyA+ RNA.
 CC The complete nucleotide sequence of isolated clones is given in
 CC Q48170.
 SQ Sequence 585 AA;

Query Match 54.4%; Score 412; DB 1; Length 585;
 Best Local Similarity 53.5%; Pred. No. 3,136-33; Indels 0; Gaps 0;
 Matches 54; Conservative 19; Mismatches 28;

Db 24 MRLADGATNOGRVEYRGQWGTICDNNVEYGVQVCRSLGPGVQAAHKAHFGCGS 83
 :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 227 IRLAGSSVHEGRVELYHAGQWGTICDDQMDADAETICQLGSLGAKMHWQAYFEEGS 266
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 84 GPWLMEVQCTGEASLADCKSLGWLKSNCRHBRDAGVCT 124
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 287 GPWLMEVQCTGEASLADCKSLGWLKSNCRHBRDAGVCT 327

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

ch_pp protein - protein database search, using Smith-Waterman algorithm
 run on: Mon Mar 13 10:25:35 2000; Maspar time 6.52 seconds
 367.019 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-147-947-6
 Description: (117-217) from US09147947A.pep (3 of 6)
 Perfect Score: 764
 Sequence: 1 VRLRGKNEFEQVEYVAVSG.....KDIWGVCPCPKMAAVTCS 101

Scoring table: PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseq35
 1:geneseqp

Statistics: Mean 28.207; Variance 100.840; scale 0.280

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	Length DB	ID	Description	Pred. No.
1	764	100.0	822	1 W99087	Human serine protease	8.43e-76
2	764	100.0	875	1 W83361	Human neurotrophin	8.43e-76
3	323	42.3	761	1 W83362	Mouse neurotrophin	8.78e-25
4	323	42.3	761	1 W99088	Mouse serine protease	8.78e-25
5	240	31.4	1785	1 W64591	Human SRCR protein	1.41e-15
6	239	31.3	489	1 W39749	Mouse macrophage protein	1.81e-15
7	239	31.3	518	1 W03561	Macrophage receptor v1	1.81e-15
8	237	31.0	451	1 R40802	Human scavenger receptor	2.99e-15
9	237	31.0	451	1 R27035	Human scavenger receptor	2.99e-15
10	237	31.0	451	1 W19708	Macrophage scavenger r	2.99e-15
11	237	31.0	451	1 W08077	Type I macrophage scav	2.99e-15
12	236	30.9	451	1 R21512	Human scavenger recept	3.84e-15
13	229	30.0	666	1 W64590	Human SRCR protein fir	2.22e-14
14	225	29.5	347	1 W68200	Human scavenger recept	6.04e-14
15	225	29.5	347	1 Y13369	Amino acid sequence of	6.04e-14
16	225	29.5	347	1 W64537	Human liver cell clone	6.04e-14
17	224	29.3	453	1 R27036	Bovine sol. scavenger	7.75e-14
18	224	29.3	453	1 R05509	Scavenger receptor pro	7.75e-14
19	219	28.7	585	1 R36533	GP85-97 clone 18 prod	2.69e-13
20	219	28.7	585	1 W81363	Human cytokine regulat	2.69e-13
21	219	28.7	585	1 R40172	Sequence of an immunor	2.69e-13
22	219	28.7	585	1 R41359	Tumour associated 90K	2.69e-13
23	219	28.7	585	1 R40213	Sequence of a 90K tumo	2.69e-13

24	217	28.4	495	1 W39747	Human macrophage recep	4.43e-13
25	217	28.4	520	1 Y04374	Human macrophage recep	4.43e-13
26	217	28.4	520	1 W39748	Human macrophage recep	4.43e-13
27	212	27.7	1290	1 W07609	Rat von Ebner's gland	1.53e-12
28	191	25.0	574	1 R40784	Sequence of cyclophilin	2.69e-10
29	191	25.0	574	1 W81362	Mouse cytokine regulat	2.69e-10
30	113	14.8	798	1 R57283	Bovine enterokinase	2.39e-02
31	89	11.6	276	1 W00104	Pimeloyl CoA synthase	4.07e+00
32	88	11.5	1130	1 R71729	Merosin major subunit	5.00e+00
33	88	11.5	1130	1 R13436	Merosin M polypeptide	5.00e+00
34	88	11.5	3110	1 R71730	Merosin major subunit	5.00e+00
35	84	11.0	349	1 W35849	Human CD5 for use in T	1.13e+01
36	83	10.9	268	1 R95967	Nucleic acid recogniti	1.38e+01
37	82	10.7	1719	1 R92100	Human RIZ allele D283	1.69e+01
38	80	10.5	113	1 W50883	Amino acid sequence of	3.07e+01
39	79	10.3	1297	1 R84639	MSOS2 protein	3.07e+01
40	79	10.3	1706	1 R92103	Rat RIZ	3.07e+01
41	79	10.3	1946	1 W47273	Lactobacillus bulgaric	3.07e+01
42	77	10.1	356	1 W46917	Amino acid sequence of	4.54e+01
43	77	10.1	1297	1 R47048	Mammalian son of seven	4.54e+01
44	75	9.8	223	1 R37573	Partial human skeletal	6.70e+01
45	75	9.8	223	1 W76805	Human ADP-ribosyltrans	6.70e+01

ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	W99087	standard: Protein; 822 AA.
AC	W99087	13-MAY-1999 (first entry)
DE	Human serine protease BSSP-3.	
DT	Serine protease; BSSP-3; brain tissue.	
KW	Homo sapiens.	
OS	MO9905290-AL.	
PN	04-FEB-1999.	
PF	24-JUL-1998; J03324.	
PR	24-JUL-1997; JP-213969.	
PA	(SUNR) SUNTOY LTD.	
PI	Tsuruoka N, Yamaguchi N, Yamashiro K;	
DR	WPI: 99-142942/12.	
DR	N-PSDB: X19024.	
PT	New serine protease expressed in brain tissue - used in screening	
PR	for potential serine protease inhibitors for drug use	
PS	Claim 1: Page 61-65; 69pp; Japanese.	
CC	The present sequence is a serine protease designated BSSP-3, which	
CC	is isolated from human brain tissue. Transformants may be used to	
CC	produce the enzyme or its partial sequences. Products from the present	
CC	invention are used for screening for potential peptide or non-peptide	
CC	serine protease inhibitors or expression regulators for use as drugs.	
SQ	Sequence 822 AA;	
Query Match	100.0%; Score 764; DB 1; Length 822;	
Best Local Similarity	100.0%; Pred. No. 8.43e-76;	
Matches	101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	117 VRLRGKNEFEQVEYVAVSGWGVTCSSHMDSDASVICHOLQGLGKIAKOTPFSGTGL 176	
Qy	117 VRLRGKNEFEQVEYVAVSGWGVTCSSHMDSDASVICHOLQGLGKIAKOTPFSGTGL 176	
Db	177 IPIYWSNVRCDENILLCERDIWQGVCPCKMAAVTCS 217	
Qy	177 IPIYWSNVRCDENILLCERDIWQGVCPCKMAAVTCS 217	
RESULT	2	
ID	W83361	standard: Protein; 875 AA.
AC	W83361	17-FEB-1999 (first entry)
DT	Human neurotrophin.	
DE	Human neurotrophin; tumour inhibition; neurological disease;	
KW	Human; neurotrophin; tumour inhibition; neurological disease;	
KW	lung disease; gene therapy; drug development; stroke; brain injury;	
KW	neurodegeneration; neuroinflammatory disease; multiple sclerosis;	
KW	epilepsy; hypoxia; ischaemia; nerve transection; neuroangiogenesis;	

KW emphysema; bronchitis.
 OS Homo sapiens.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PE 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72589.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1; Page 20-24; 50pp; English.
 CC The present sequence represents human neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis)
 SO Sequence 875 AA:

Query Match 100.0%; Score 764; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 8.43e-76;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 170 VRLRGKNEFEETVEYASGVGWTCSHWDSDASVICHOLGKGAKIKPFSGGLG 229
 QY 117 VRLRGKNEFEETVEYASGVGWTCSHWDSDASVICHOLGKGAKIKPFSGGLG 176
 Db 230 IPIYWSNVRGDEENILCEKIDIMOGVCPOKMAAVTC 270
 QY 177 IPIYWSNVRGDEENILCEKIDIMOGVCPOKMAAVTC 217

RESULT 3
 ID W83362 standard; Protein; 761 AA.
 AC W83362;
 PT 17-FEB-1999 (first entry)
 PI Mouse neurotysin.

Mus: neurotysin; tumour inhibition; neurological disease;
 Lung disease; gene therapy; drug development; stroke; brain injury;
 neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 epilepsy; hypoxia; ischaemia; nerve transection; neovascularisation;
 emphysema; bronchitis.

OS Mus musculus.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PE 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72590.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1; Page 29-32; 50pp; English.
 CC The present sequence represents mouse neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SO Sequence 761 AA:

Query Match 42.3%; Score 323; DB 1; Length 761;
 Best Local Similarity 45.5%; Pred. No. 8.75e-25;
 Matches 46; Conservative 19; Mismatches 34; Indels 2; Gaps 2;

Db 166 IRLVGNSGHEGRVELYHAGWGITICDDQMNADADVICRQLGSLGAKAHQAHF-GE 224
 QY 117 VRLRGKNEFEETVEYASGVGWTCSHWDSDASVICHOLGKGAKIA-KQPFSGGLG 175
 Db 225 SGPIILDEVRCGTNELSTFQCPKSSMGHNCGRKEDAVSC 265
 QY 176 IPIYWSNVRGDEENILCEKIDIMOGVCPOKMAAVTC 216

RESULT 4
 ID W99088 standard; Protein; 761 AA.
 AC W99088;
 DT 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 KW Serine protease; BSSP-3; Brain tissue.
 OS Mus sp.
 PN W09905290-A1.
 PD 04-FEB-1999.
 PE 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SUNR) SONTORY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR WPI: 99-142942/12.
 DR N-PSDB: X19027.

PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1; Page 51-54; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transformants may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SO Sequence 761 AA:

Query Match 42.3%; Score 323; DB 1; Length 761;
 Best Local Similarity 45.5%; Pred. No. 8.75e-25;
 Matches 46; Conservative 19; Mismatches 34; Indels 2; Gaps 2;

Db 166 IRLVGNSGHEGRVELYHAGWGITICDDQMNADADVICRQLGSLGAKAHQAHF-GE 224
 QY 117 VRLRGKNEFEETVEYASGVGWTCSHWDSDASVICHOLGKGAKIA-KQPFSGGLG 175
 Db 225 SGPIILDEVRCGTNELSTFQCPKSSMGHNCGRKEDAVSC 265
 QY 176 IPIYWSNVRGDEENILCEKIDIMOGVCPOKMAAVTC 216

RESULT 5
 ID W64591 standard; Protein; 1785 AA.
 AC W64591;
 DT 23-OCT-1998 (first entry)
 DE Human SRCR protein.
 KW Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;
 KW nervous system; medullo-blastoma; glioma; breast; detection;
 KW autoantibody; ss.
 OS Homo sapiens.

[illegible]

Best Local Similarity 39.28; Pred. No. 1,81e-15;
Matches 40; Conservative 21; Mismatches 34; Indels 7; Gaps 6;

Db 423 VRLMGSTNR--GRAEYVNNMGITICDDMDNPNATVCRMLGXS-RCRA-LSYVG-GS 477
117 VRLRGKNEEGYEVYASGVWGVCSHMDSDASVICHOLQJLG-KGIAKOTPSGIGL 176

Db 478 GNI-WLDNVNCRGTENSLMDCSKNSWGNHNCVHNEDAGVECS 518
177 LPIYV-SNVRCRGDEENILLCEKIDWOGVCPQKMAAAVTCs 217

RESULT 8
ID R40802 standard; Protein; 451 AA.
AC R40802;
DT 23-FEB-1994 (first entry)
DE Human scavenger receptor I.
Human: scavenger receptor; antihuman scavenger receptor antibody;
AHSR: mammal; rabbit; antiserum, monoclonal; HAT-resistant hybridoma;
arteriosclerosis.
Homo sapiens.
J05192179-A.
03-AUG-1993.
16-AUG-1991: 229728.
27-AUG-1990: JP-222398.
PA (CHUS) CHUGAI PHARM CO LTD.
DR WPI: 93-277488/35.
DR N-PSDB: Q47730.
PT Anti-human scavenger receptor antibody - useful for monitoring
progress of arteriosclerosis
PS Claim 2: Page 5-7; 10pp; Japanese.
CC The sequences given in R40802-03 represent the human scavenger
CC receptors I and II. These proteins may be used in the production
CC of antihuman scavenger receptor antibodies (AHSR). These peptides
CC were used to immunise a mammal, pref. a rabbit, and anti-serum was
CC prepared. A monoclonal antibody was prepared from HAT-resistant
CC hybridoma. These antibodies may be used to determine the progress
CC of arteriosclerosis.
SQ Sequence 451 AA:

Query Match 31.0%; Score 237; DB 1; Length 451;
Best Local Similarity 37.9%; Pred. No. 2.99e-15;
Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

Db 350 VRLVGSQPHRGREYILHSGOMGTICDDREVRVGVVCRSLGYPGVQVAKAAHF-GQG 408
117 VRLRGKNEEGYEVYASGVWGVCSHMDSDASVICHOLQJLG-KGIAKOTPSGIGL 175

QY 409 TGPI-WLNEVFCGREGSIEECKIRQWGTACSHSDAGVTCT 450
176 LPIYV-SNVRCRGDEENILLCEKIDWOGVCPQKMAAAVTCs 217

RESULT 9
ID R27035 standard; Protein; 451 AA.
AC R27035;
DT 16-FEB-1993 (first entry)
DE Human sol. scavenger receptor.
KM Macrophage: endotoxaemia; radiolabelled; toxic; degenerate.
OS Homo sapiens.
FH key
FT Location/Qualifiers
FT 1..50
FT /note= "cytoplasmic"
FT 51..76
FT /note= "transmembrane"
FT 77..109
FT /note= "spacer"
FT 110..271
FT /note= "alpha helical coiled coil"
FT 272..343
FT /note= "collagen binding"
FT 341..451
FT /note= "Cys-rich"

PN W09214482-A.
PD 03-SEP-1992.
PF 21-FEB-1992; 001370.
PR 22-FEB-1991; US-662227.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Krieger M.
DR WPI: 92-315935/38.
DR N-PSDB: Q28539.
PT Treatment and diagnosis of endotoxaemia and related disease
PT states - using a polypeptide fragment of the extracellular
PT portion of a macrophage scavenger receptor protein
PS Disclosure; Page 33; 56pp; English.
CC The human scavenger receptor protein sequence was deduced from the
CC DNA sequence obtd. by screening a platelet cDNA library with
CC degenerate probes designed based on fragments of the purified
CC scavenger receptor protein. The scavenger receptor protein has a
CC binding capacity for acetylated low density lipoprotein (LDL) and
CC binds endotoxin. Fragments of the extracellular portion of the
CC scavenger receptor protein inactivate endotoxin-related substances
CC and are used in the diagnosis and treatment of endotoxaemia. They
CC may be administered to patients at high risk of symptomatic or
CC endotoxic shock. The fragments may also be fixed to inert supports
CC for purification purposes. The fragments may be modified to have
CC greater binding affinity for the endotoxin-related substance than
CC the native scavenger receptor protein, or to more effectively
CC neutralise the toxic or pathogenic effects of mols. that bind
CC the scavenger receptor protein or of organisms which express such mols.
CC See also R27036.
SQ Sequence 451 AA:

Query Match 31.0%; Score 237; DB 1; Length 451;
Best Local Similarity 37.9%; Pred. No. 2.99e-15;
Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

Db 350 VRLVGSQPHRGREYILHSGOMGTICDDREVRVGVVCRSLGYPGVQVAKAAHF-GQG 408
117 VRLRGKNEEGYEVYASGVWGVCSHMDSDASVICHOLQJLG-KGIAKOTPSGIGL 175

QY 409 TGPI-WLNEVFCGREGSIEECKIRQWGTACSHSDAGVTCT 450
176 LPIYV-SNVRCRGDEENILLCEKIDWOGVCPQKMAAAVTCs 217

RESULT 10
ID W19708 standard; Protein; 451 AA.
AC W19708;
DT 19-AUG-1997 (first entry)
DE Macrophage scavenger receptor protein.
KM Macrophage scavenger receptor protein; MSRP; human; lipoteichoic acid;
KM LTA; Gram-positive bacteria; cell wall; arteriosclerosis; host defence;
KM septiccaemia; inhibitor; complement activation; cytokine release; therapy;
KM nitric oxide production; bacterial infection; septic shock.
OS Homo sapiens.
PN US5624904-A.
PD 29-APR-1997.
PF 17-NOV-1993; 154365.
PR 17-NOV-1993; US-154365.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI (UYVA) UNIT VALE.
PI Uolner KA, Krieger M.
DR WPI: 97-258236/23.
DR N-PSDB: T68796.
PT Treatment of septiccaemia caused by Gram-positive bacteria - by
PT administration of macrophage scavenger receptor protein
PS Disclosure; Column 23-26; 16pp; English.
CC This sequence represents the human macrophage scavenger receptor protein
CC (MSRP). MSRP binds specifically to the lipoteichoic acid (LTA) residues
CC on the Gram-positive bacterial cell wall. MSRP exhibit unusually broad
CC binding specificity for polyanionic ligands, and have been implicated in
CC arteriosclerosis and a variety of host defence functions. This protein can
CC be used in the method of the invention. The method of the invention is
CC for the treatment of septiccaemia caused by Gram-positive bacteria. The
CC method comprises administration of a MSRP in an amount sufficient to

CC inhibit complement activation, cytokine release or nitric oxide
 CC production induced by LTA released by the Gram-positive bacteria. In
 CC addition to treating Gram-positive septic shock, the MSRP can be used to
 CC screen for other compounds for treating Gram-positive septic shock. MSRP
 CC can also be used to purify, label or detect LTA or LTA-containing cells.
 CC MSRP, active MSRP fragments, anti-MSRP antibodies or other compounds that
 CC inhibit binding of Gram-positive bacteria to MSRP can be used to treat
 CC pathologies such as septicemia, Gram-positive bacterial infection,
 CC Gram-positive septicemia or Gram-positive or Gram-negative septic shock.
 SQ Sequence 451 AA:

Query Match 31.0%; Score 237; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 2,99e-15;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

350 VRLVGGSGPHEGRVILHSGOMGTICDDNREVRVGVCRSLGPGVQVHKAHF-GOG 408
 117 VRLRGKNEFEGLVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175
 409 TGPI-WLNEVFCGRSSIECKIRMGTRACSHSDAGVTCT 450
 176 LPIIYW-SNVRCRGDEENILCEKIDIMOGVCPOKMAAAYTCS 217

RESULT 11
 ID W08077 standard; Protein: 451 AA.

AC W08077;
 DT 25-FEB-1997 (first entry)
 DE Type I macrophage scavenger receptor.
 KM Type I macrophage scavenger receptor; SRG; kidney; 293 cell;
 OS Homo sapiens.
 PN W09638725-A1.
 PD 05-DEC-1996.
 PE 30-MAY-1996; 008081.
 PR 30-MAY-1995; US-453117.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI Branner ME, Elshourbagy NA, Lysko PG;
 DR N-PSDB; T43259.
 PT New modified human embryonic kidney 293 cells - transfected with a
 PT mammalian scavenger receptor gene to enhance ability to attach to a
 PT solid support
 PS Disclosure: Page 10-13; 32pp; English.
 CC Human type I (W08077) and type II (W08078) macrophage scavenger
 CC receptors can be expressed in human embryonic kidney 293 cells
 CC following transfection of the cells with vectors carrying the
 CC respective scavenger receptor genes (T43259 and T43260). The
 CC transfected cells show an enhanced ability to attach to a solid
 CC support and are useful for screening cpds. for biological activity,
 CC or for identifying antagonists of the scavenger receptor gene, e.g.
 CC to develop agents for treatment of atherosclerosis.
 SQ Sequence 451 AA:

Query Match 31.0%; Score 237; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 2,99e-15;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHEGRVILHSGOMGTICDDNREVRVGVCRSLGPGVQVHKAHF-GOG 408
 117 VRLRGKNEFEGLVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175
 DB 409 TGPI-WLNEVFCGRSSIECKIRMGTRACSHSDAGVTCT 450
 176 LPIIYW-SNVRCRGDEENILCEKIDIMOGVCPOKMAAAYTCS 217

RESULT 12
 ID R21512 standard; Protein: 451 AA.
 AC R21512;
 DT 19-MAY-1992 (first entry)
 DE Human scavenger receptor type I.
 KW Lipoproteins.

OS Homo sapiens.
 PN J03290184-A.
 PD 19-DEC-1991.
 PE 06-APR-1990; 090274.
 PR 06-APR-1990; JP-090274.
 PA (CHUS) CHUGAI PHARMACEUTICAL KK.
 DR WPI: 92-051436/07.
 DR N-PSDB; Q21549.
 PT New scavenger receptor-producing animal cells - which have been
 PT transformed with vector contg. gene for coding human scavenger
 PT receptor I or II type under control of promoter, etc.
 PS Disclosure: Fig 1; 3pp; Japanese.
 CC The gene encoding the receptor can be expressed from a cytomagalo-
 CC virus promoter in a host cell e.g. CHO cell. The resulting
 CC recombinant scavenger receptor can be used to detect modified lipo-
 CC proteins or modified substances in the blood.
 CC See also R21513.
 SQ Sequence 451 AA:

Query Match 30.9%; Score 236; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 3,84e-15;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHEGRVILHSGOMGTICDDNREVRVGVCRSLGPGVQVHKAHF-GOG 408
 117 VRLRGKNEFEGLVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175
 DB 409 TGPI-WLNEVFCGRSSIECKIRMGTRACSHSDAGVTCT 450
 176 LPIIYW-SNVRCRGDEENILCEKIDIMOGVCPOKMAAAYTCS 217

RESULT 13

ID W64590 standard; Protein: 666 AA.
 AC W64590;
 DT 23-OCT-1998 (first entry)
 DE Human SRCR protein fragment.
 KM Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;
 KW nervous system; medullo-blastoma; glioma; breast; detection;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 3
 FT Key Location/Qualifiers
 FT Misc_difference 3 /label= unknown
 PN W09830687-A2.
 PD 16-JUN-1998.
 PE 09-JAN-1998; D00096.
 PR 18-JUL-1997; DE-030997.
 PR 09-JAN-1997; DE-000519.
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 PI Mollenhauer J, Poustka A;
 DR WPI: 98-399136/34.
 DR N-PSDB; V49651.
 PT Proteins containing scavenger receptor, cysteine rich domain -
 PT useful for diagnosis and treatment of tumours
 PS Claim 1; Fig 1; 54pp; German.
 CC This sequence represents a fragment of a human protein which contains a
 CC SRCR (scavenger receptor, cysteine-rich) domain. The gene and encoded
 CC protein can be used to diagnose or treat tumours, particularly of the
 CC nervous system (medullo-blastoma or glioma) or breast. The DNA sequence
 CC and probes derived from it, are used to identify genes that express
 CC SRCR domain containing proteins, to determine the form in which these
 CC proteins exist and to assess the significance of individual forms on
 CC cellular properties. The protein can be used to detect the presence of
 CC autoantibodies, and Ab which regulate its expression.
 SQ Sequence 666 AA:

Query Match 30.0%; Score 229; DB 1; Length 666;
 Best Local Similarity 38.2%; Pred. No. 2,22e-14;
 Matches 39; Conservative 16; Mismatches 45; Indels 2; Gaps 2;

DB 61 LRLVGGDRCRVRYVLYRGSGVGYCDSDWTNDANVYCRQIGCGWASAPGNARF-GOG 119
 117 VRLRGKNEFEGLVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175

CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to growth
 CC or survival of nerve cells including Parkinson's disease, Alzheimer's
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 SQ Sequence 347 AA;

Query Match 29.5% Score 225; DB 1; Length 347;
 Best Local Similarity 35.6%; Pred. NO. 6.04e-14;
 Matches 37; Conservative 20; Mismatches 43; Indels 4; Gaps 4;
 DB 244 LRLVGGDNLCSGRLEVLHKGWGSVCDDNMGEKEDQVCKQLGCGKSLSPFRDRKCYGP 303
 QY 117 VRLRGKNEFEGRVEYVAYASGVWGTVCSSHWDDSDASVICHQLDLG-GKGLA-KQTFPSGL 174
 DB 304 GVGRI-WLDNVRCSGEGQSLDQCQHRFWGPHDCTHOEDNAVICS 346
 QY 175 GLPIYW-SNVRGRGDENTILCEKIDWGGVCPQKMAAVTCS 217

Search completed: Mon Mar 13 10:25:46 2000
 Job time : 11 secs.

THIS PAGE BLANK (OPTION)

334-433/SCRC3
Use last No. 5

WIDE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h-pe protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:31:53 2000; Maspar time 6.67 Seconds
Tabular output not generated. 355.116 Million cell updates/sec

Title: >US-09-147-947-6
Description: (334-433) from US09147947A.pep (5 of 6)
Perfect Score: 787
Sequence: 1 IRLAGGSGHSEGRLEVVYRG.....RRQGRHDCSHREDVSACY 100

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 28.512; Variance 102.112; scale 0.279

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
787	100.0	822	1 W99087	Human serine protease	3.08e-78
787	100.0	875	1 W83361	Human neurotysin.	5.48e-67
691	87.8	761	1 W83362	Mouse neurotysin.	5.48e-67
691	87.8	761	1 W99088	Mouse serine protease	5.48e-67
412	52.4	451	1 W19708	Macrophage scavenger r	1.05e-34
407	51.7	451	1 R27035	Human sol. scavenger r	3.88e-34
407	51.7	451	1 R40802	Human scavenger recept	3.88e-34
407	51.7	451	1 W08077	Type I macrophage scav	3.88e-34
406	51.6	451	1 R2512	Human scavenger recept	5.05e-34
404	51.3	453	1 R05509	Scavenger receptor pro	8.53e-34
10	404	51.3	1 R27036	Bovine sol. scavenger	8.53e-34
11	404	51.3	1 R40172	Sequence of an immunor	3.82e-29
12	363	46.1	1 R40213	Gp85-97 clone 18 prod.	3.82e-29
13	363	46.1	1 W81363	Human cytokine regulat	3.82e-29
14	363	46.1	1 W81363	Tumour associated 90K	4.96e-29
15	363	46.1	1 W64590	Human SRCR protein fra	4.96e-29
16	362	46.0	1 W64591	Mouse SRCR protein.	6.67e-28
17	352	44.7	1 W39749	Macrophage receptor wi	6.67e-28
18	352	44.7	1 W03561	Rat von Ebner's gland	8.65e-28
19	351	44.6	1 W39747	Human macrophage recep	2.44e-27
20	347	44.1	1 Y04374		

24	347	44.1	520	1 W39748	Human macrophage recep	2.44e-27
25	343	43.6	347	1 W64537	Human liver cell clone	6.89e-27
26	343	43.6	347	1 W68200	Human scavenger recept	6.89e-27
27	343	43.6	347	1 Y13369	Amino acid sequence of	1.94e-26
28	339	43.1	574	1 R40784	Sequence of cyclophilin	1.94e-26
29	339	43.1	574	1 W81362	Mouse cytokine regulat	1.59e-01
30	105	13.3	349	1 W35849	Human CD5 for use in T	1.59e-01
31	96	12.2	356	1 W46917	Amino acid sequence of	1.08e+00
32	96	12.2	798	1 R57283	Bovine enterokinase.	1.08e+00
33	84	10.7	2509	1 W32881	Protein (OA-519) cross	2.31e+01
34	81	10.3	135	1 R28982	Thyroid N-acetyl-glucos	4.16e+01
35	78	9.9	337	1 W06579	Lipo-oligosaccharide g	4.16e+01
36	78	9.9	337	1 R91314	N. gonorrhoeae glycosyl	4.16e+01
37	78	9.9	872	1 W75912	Helicobacter valyl-L-TRN	5.06e+01
38	77	9.8	172	1 W72084	HSV-2 strain SB5 Conti	5.06e+01
39	77	9.8	172	1 W72210	HSV-2 strain SB5 Conti	5.06e+01
40	77	9.8	193	1 W72157	HSV-2 strain SB5 Conti	5.06e+01
41	77	9.8	335	1 R25063	Soluble human IL-5 rec	5.06e+01
42	77	9.8	420	1 R22215	Sequence of human inte	5.06e+01
43	77	9.8	420	1 R22219	Sequence of secretory	5.06e+01
44	77	9.8	874	1 W98698	H. pylori GHPD 686 pro	5.06e+01
45	77	9.8	1250	1 W59361	Drosophila melanogaste	5.06e+01

ALIGNMENTS

RESULT 1
ID W99087 standard; Protein: 822 AA.
AC W99087;
DT 13-MAR-1999 (first entry)
DE Human serine protease BSSP-3.
KW Serine protease; BSSP-3; brain tissue.
OS Homo sapiens.
PN W09905290-A1.
PF 04-FEB-1999.
PF 24-JUL-1998; J03324.
PR 24-JUL-1997; JP-213699.
PA (SUNR) SUNTOXY LTD.
PI Tsunokuchi N, Yamaguchi N, Yamashiro K;
DR WPI: 99-142942/12.
DR N-PSDB; X19024.
PT New serine protease expressed in brain tissue - used in screening
PT for potential serine protease inhibitors for drug use
PS Claim 1: Page 61-65; 69pp; Japanese.
CC The present sequence is a serine protease designated BSSP-3, which
CC is isolated from human brain tissue. Transformants may be used to
CC produce the enzyme or its partial sequences. Products from the present
CC invention are used for screening for potential peptide or non-peptide
CC serine protease inhibitors or expression regulators for use as drugs.
SQ Sequence 822 AA;

Query Match 100.0%; Score 787; DB 1; Length 822;
Best Local Similarity 100.0%; Pred. No. 3.08e-78;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 334 IRLAGGSGHSEGRLEVVYRGWGTCDGWTETLVVVCROLGKRYKXASNFESTG 393
334 IRLAGGSGHSEGRLEVVYRGWGTCDGWTETLVVVCROLGKRYKXASNFESTG 393
QY 394 PIVLDVSCSGKETRFLOCSRRQGRHDCSHREDVSACY 433
394 PIVLDVSCSGKETRFLOCSRRQGRHDCSHREDVSACY 433
RESULT 2
ID W83361 standard; Protein: 875 AA.
AC W83361;
DT 17-FEB-1999 (first entry)
DE Human neurotysin.
KW Human; neurotysin; tumour inhibition; neurological disease;
KW lung disease; gene therapy; drug development; stroke; brain injury;
KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;
KW epilepsy; hypoxia; ischaemia; nerve transection; neovascularogenesis;

KW emptysema; bronchitis.
OS Homo sapiens.
PN WO9849322-A1.
PD 05-NOV-1998.
PT 24-APR-1998; 1B0625.
PR 26-APR-1997; CH-000966.
PA (SOND/) SONDEREGGER P.
PI Sonderegger P;
DR WPI: 99-009438/01.
DR N-PSDB: V72589.
PT New human and murine neurotysin - used, e.g. for inhibiting
PT tumours, treatment of neurological or lung disease, including by
PT gene therapy and in drug development
PS Claim 1: Page 20-24; 50pp: English.
CC The present sequence represents human neurotysin. Neurotysin proteins
CC and polynucleotides can be used: (i) to inhibit tumours, including
CC metastases; e.g. of brain or retina; (ii) to minimise tissue damage
CC caused by stroke or brain injury (having a protective effect on the
CC penumbra zone); (iii) to treat or prevent neurodegeneration,
CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
CC increase survival of damaged neurons (e.g. in cases of hypoxia,
CC ischaemia, nerve transection) and to stimulate regeneration and/or
CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
CC degeneration or neovascularogenesis); (vi) to prevent apoptosis (or other
CC causes of cell death) in the nervous system; (vii) to regenerate brain
CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
CC performance, including learning and memory; (x) to treat or prevent a
CC wide range of psychiatric disorders; and (xi) to treat brain or lung
CC injury associated with protease expression (specifically emptysema or
CC bronchitis).
SQ Sequence 875 AA.

Query Match	100.0%;	Score 787;	DB 1;	Length 875;
Best Local Similarity	100.0%;	Pred. No. 3.08e-78;		
Matches	100;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Db	387	IRGGGKSHSGRLRELVYVYRGQMGVCDGDMGELMTNYVVCRLGKRYKQKQSANFEESTG	446
Qy	334	IRLQGGKSHSGRLRELVYVYRGQMGVCDGDMGELMTNYVVCRLGKRYKQKQSANFEESTG	353
Db	447	PIWLDVSCSGKETRFLOCSRRMGHDCSHREDVSTACY	486
Qy	394	PIWLDVSCSGKETRFLOCSRRMGHDCSHREDVSTACY	433

RESULT	3
ID	W83362 standard; Protein; 761 AA
AC	W83362;
DT	17-FEB-1999 (first entry)

Mouse neurotropsin.

Mouse; neurotrophic; tumour inhibition; neurological disease; lung disease; gene therapy; drug development; stroke; brain injury; neurodegeneration; neuroinflammatory disease; multiple sclerosis; emphysema; hypoxia; ischaemia; nerve transection; neuroangiogenesis; epilepsy; bronchitis.

OS	Mus musculus.
PN	WO9849322-A1.

PD 05-NOV-1998.
PF ~~24-APR-1998;~~

PR 26-APR-1997; CH-000966
PA (SOND/) SONDEREGGER P.

PI Sonderegger P;
DR WPI; 99-009438/01.

DR	N-PSDB; V72590
PT	New human and r

PT tumours, treatment of neurological o
gene therapy and in drug development

CC The present sequence represents mouse neurotrophin proteins

CC and polynucleotides can be used: (1) to inhibit tumours, including metastases, e.g. of brain or retina; (11) to minimise tissue damage

CC caused by stroke or brain injury (having a protective effect
CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
CC increase survival of damaged neurons (e.g. in cases of hypoxia,
CC ischaemia, nerve transection) and to stimulate regeneration and/or
CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
CC degeneration or neovascularogenesis); (vi) to prevent apoptosis (or other
CC causes of cell death) in the nervous system; (vii) to regenerate brain
CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
CC performance, including learning and memory; (x) to treat or prevent a
CC wide range of psychiatric disorders; and (xi) to treat brain or lung
CC injury associated with protease expression (specifically emphysema or
CC bronchitis). /61 AA; .

Query Match	87.8%;	Score 691;	DB 1;	Length 761;
Best Local Similarity	82.0%;	Pred. No. 5.48e-67;		
Matches	82;	Conservative	10;	Mismatches 8; Indels 0; Gaps 0

```

Db      273  IRLAGGKSTHEGRLEYYIKGQGWTCDDGWTENMTYVACLLGFKYKQSSVNHFDGSNR 332
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      334  IRLAGGKSGHEGRLEYYIRGQGWTCDDGWTENMTYVVCRLQDGFKKGQASNHFEESTG 393

```

Db 333 PVLDDVSCSGEVSFTQCSRRQMGKHC SHRDVGLTCY 372
 ||||| : ||||| : : : : :
 QY 394 PVLDDVSCSGEVSFTQCSRRQMGKHC SHRDVSLACY 433

RESULT	4
ID	W99088 standard; protein; 761 AA

AC W99088;
DT 13-MAY-1999 (first entry)

DE Mouse serine protease BSSP-3.
KW Serine protease; BSSP-3; brain tissue

OS	Mus sp.
PN	W09905290-A1

PD 04-FEB-1999.
PF 24-JUL-1998; J03324.
24 JUL 1997

PR 24-JUL-1997; JP-213969.
PA (SUNR) SUNTORY LTD.
Z

PI TSURUOKA N, Yamaguchi
DR WPI; 99-142942/12.

DR N-PSDB; A1902/.

PT New serine protease expressed in brain tissue - used in screening

For potential serine protease inhibitor
Example 1; Page 51-54; 69pp; Japanese

CC The present sequence is a serine protease designated B5B-3, which
CC is isolated from mouse brain tissue. Transformants may be used to

CC produce the enzyme or its partial sequences. Products from the present
CC invention are used for screening for potential peptide or non-peptide

CC serine protease inhibitors or expression regulators for use as drugs.
SQ Sequence 761 AA;

Query Match	87.8%	Score 691	DB 1	Length 761
-------------	-------	-----------	------	------------

Best Local Similarity 82.0%; Pred. NO. 5.48e-6/;
Matches 82; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 273 IRLAGKSTHGRLEVYKQMGTVCDGWTENMTYACRLLGPRYKQSSVNHFDGSR 332

0Y 334 IRLAGKSGSHEGRLEVVYRQMGWTCDDGWTELMTTVVVCROUGFKYKQASANHPFEESTG 3933

Db 333 P IWL DVSCGKEVSFIQCSRQWGRHDCSHREDVGLTCY 372

QY 394 P I W L D V S C G K E T R F L Q C S R Q W G R H D C S H R E D V S I A C Y 433

1
2
3
4
5
6
7

RESULT	5
ID	W19708 standard; Protein; 451 AA.

AC W19/08;
DT 19-AUG-1997 (first entry)

DE Macrophage scavenger receptor protein.
KW Macrophage scavenger receptor protein; MSRP; human; lipoteichoic acid;

KW septicaemia; inhibitor; complement activation; cytokine release; therapy.

kw nitric oxide production; bacterial infection; septic shock

OY 392 TGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 8

ID W08077: standard; Protein: 451 AA.

AC W08077:

DT 25-FEB-1997 (first entry)

DE Type I macrophage scavenger receptor. SRG; kidney; 293 cell;

KW Type I macrophage scavenger receptor. SRG; kidney; 293 cell;

OS Homo sapiens. cell attachment; atherosclerosis.

PN W09638725-A1.

PD 05-DEC-1996.

PF 30-MAY-1996; U08081.

PR 30-MAY-1995; US-453117.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PI Brawner ME, Elshourbady MAE, Lysko PG;

DR WPI: 97-034505/03.

PN N-PSDB: T43259.

DE New modified human embryonic kidney 293 cells - transfected with a

human scavenger receptor gene to enhance ability to attach to a

solid support.

PS Disclosure: Page 10-13; 32pp; English.

CC Human type I (W08077) and type II (W08078) macrophage scavenger

receptors can be expressed in human embryonic kidney 293 cells

following transfection of the cells with vectors carrying the

respective scavenger receptor genes (T43259 and T43260). The

transfected cells show an enhanced ability to attach to a solid

support and are useful for screening cpds. for biological activity,

or for identifying antagonists of the scavenger receptor gene, e.g.

to develop agents for treatment of atherosclerosis.

Sequence 451 AA.

Query Match 51.7%; Score 407; DB 1; Length 451;

Best Local Similarity 51.5%; Pred. No. 3.88e-34;

Matches 52; Conservative 18; Mismatches 28; Indels 3; Gaps 2;

DB 350 VRLVGGSGPHEGVEILHSGOMGTICDDREVRVGVGCRSLGTP-GVQAVHKAHFGOG 408

Y 334 IRLAGKGSGHEGRLEYYRGOMGTVCDDGTELTNTYVCRQLGFKYKQA--SANHEERS 391

DB 409 TGPIMLNEVFCFRESSIECKRIRMGTRACSHSDAGVTC 449

Y 392 TGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 9

ID R21512: standard; Protein: 451 AA.

AC R21512:

DT 19-MAY-1992 (first entry)

DE Human scavenger receptor type I.

OS Homo sapiens.

PN J03290184-A.

PD 19-DEC-1991.

PF 06-APR-1990; 090274.

PR 06-APR-1990; JP-090274.

PA (CHUS) CHUGAI PHARMACEUTICAL KK.

DR WPI: 92-051436/07.

DR N-PSDB: 021549.

PT New scavenger receptor-producing animal cells - which have been

transformed with vector contg. gene for coding human scavenger

receptor I or II type under control of promoter, etc.

PS Disclosure: Fig 1; 9pp; Japanese.

CC The gene encoding the receptor can be expressed from a cytomegalo-

virus promoter in a host cell e.g. CHO cell. The resulting

recombinant scavenger receptor can be used to detect modified lipo-

proteins or modified substances in the blood.

CC See also R21513.

Sequence 451 AA;

Query Match 51.6%; Score 406; DB 1; Length 451;

Best Local Similarity 50.5%; Pred. No. 5.05e-34;

Matches 51; Conservative 19; Mismatches 28; Indels 3; Gaps 2;

RESULT 10

ID R05509: standard; protein: 453 AA.

AC R05509;

DT 23-OCT-1990 (first entry)

DE Scavenger receptor protein with affinity for acylated low density

lipoprotein (LDL).

KW Acylated low density lipoprotein; LDL; atherosclerotic plaque;

ds

PN W09005748-A.

PD 31-MAY-1990.

PF 14-NOV-1989; 005116.

PR 15-NOV-1988; US-272002.

PA (MAST) Massachusetts Inst Tech.

PI Engellennet TT.

DR WPI: 90-193408/25.

DR N-PSDB: 004926.

PT New receptor protein -

has affinity for acetylated low density lipoprotein and

corresponding antibodies and DNA sequences.

PS Disclosure: 79pp; English.

CC Receptor protein, and fragments thereof may be

immobilised on a support and used in assay and purification of the

CC LDL target. Labelled Abs, raised to the protein may be injected

into the vascular system to detect the presence of atherosclerotic

CC plaques.

Sequence 453 AA.

Query Match 51.3%; Score 404; DB 1; Length 453;

Best Local Similarity 52.0%; Pred. No. 8.53e-34;

Matches 53; Conservative 19; Mismatches 25; Indels 5; Gaps 4;

DB 352 VRLVGGSGPHEGVEILHSGOMGTVCDDRW-ELRGVLVCRSLGK-GVQSHKAAHYGK 409

Y 334 IRLAGKGSGHEGRLEYYRGOMGTVCDDGTEL-NTYVCRQLGFKYKQASAH--FEE 390

DB 410 TGPIMLNEVFCFRESSIECKRIRMGTRACSHSDAGVTC 451

Y 391 TGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 11

ID R27036: standard; Protein: 453 AA.

AC R27036;

DT 16-FEB-1993 (first entry)

DE Bovine sol. scavenger receptor.

KW Macrophage; endotoxaemia; radiolabelled; toxic; degenerate.

OS Bos taurus.

DR Key

FT Location/Qualifiers

FT 1..50

FT /note="cytoplasmic"

FT 51..76

FT /note="transmembrane"

FT 77..109

FT /note="spacer"

FT 110..271

FT /note="alpha helical coiled coil"

FT 272..343

FT /note="collagen binding"

FT 341..451

FT /note="Cys-rich"

PN W09214482-A.

PD 03-SEP-1992.

PT Recombinant DNA - encoding tumour associated antigen,
 PT immuno-regulin-95
 PS Disclosure: Fig 1: 69pp: English.
 CC Immunoregulin-95 is a 90k tumour-associated antigen purified from
 CC the culture fluid of the human breast cancer cell line, CG-5, the
 CC serum of a breast cancer patient, or the ascitic fluid from an
 CC ovarian cancer patient. The native antigen, which has a mol. wt. of

D**b** 24 MRLADGATNCGREIHYRCGMVCONLDDLDASVCALGEENATOAGRAFGQS 83
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 334 IRLAGGGSHSGREIVYYRGMGVCDGTENTNYTQCQLDFKXYGKA - SANHEEST 392
D**b** 84 GPIMLDEVCTGTASTLADCKSLGWLNCRHRRDAGVYC 123
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Thu Mar 16 07:57:59 2000

US-09-147-947-6-05.ra1

Page 1

384-433/SCRC3

4th Mo. / 15 redundant with 1st Mo. 8th A. Gene Seq

W O E N E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:32:18 2000; MasPar time 4.66 Seconds
Tabular output not generated. 278.133 Million cell updates/sec

Title: >US-09-147-947-6
Description: (334-433) from US09147947A.pep (5 of 6)
Sequence: 1 IRLAGKSGSHGRLEVVYRG.....RROMGRHDCSHREDVSACY 100

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfile1

Statistics: Mean 27.056; Variance 104.512; scale 0.259

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length DB	ID	Description	Pred. No.
412	52.4	451	1	US-08-154-Sequence 2, Applicatio	1.37e-32
407	51.7	451	2	US-08-973-Sequence 2, Applicatio	4.65e-32
407	51.7	451	1	US-08-453-Sequence 2, Applicatio	4.65e-32
407	51.7	451	2	US-08-948-Sequence 2, Applicatio	4.65e-32
407	51.7	451	3	PCT-US96-0Sequence 2, Applicatio	4.65e-32
404	51.3	453	4	5510466-4Patent No. 5510466.	9.67e-32
7	363	46.1	585	1 US-08-473-Sequence 10, Applicati	2.08e-27
8	363	46.1	585	2 US-08-316-Sequence 10, Applicati	2.08e-27
9	363	46.1	585	1 US-08-477-Sequence 10, Applicati	2.08e-27
10	362	44.7	489	2 US-08-794-Sequence 7, Applicatio	2.99e-26
11	362	44.7	518	1 US-08-392-Sequence 2, Applicatio	2.99e-26
12	351	44.6	1290	1 US-08-470-Sequence 2, Applicatio	3.81e-26
13	347	44.1	495	2 US-08-794-Sequence 6, Applicatio	1.00e-25
14	347	44.1	520	2 US-08-794-Sequence 6, Applicatio	1.00e-25
15	96	12.2	356	2 US-08-681-Sequence 1, Applicatio	9.00e-01
16	96	12.2	798	3 PCT-US94-0Sequence 2, Applicatio	9.00e-01
17	96	12.2	798	1 US-08-200-Sequence 2, Applicatio	9.00e-01
18	84	10.7	2509	1 US-08-469-Sequence 10, Applicati	9.39e+00
19	81	10.3	23	2 US-08-316-Sequence 11, Applicati	1.66e+01
20	81	10.3	23	1 US-08-477-Sequence 11, Applicati	1.66e+01
21	81	10.3	23	1 US-08-473-Sequence 12, Applicati	2.92e+01
22	78	9.9	337	1 US-08-683-Sequence 5, Applicatio	2.92e+01
23	78	9.9	337	1 US-08-683-Sequence 5, Applicatio	2.92e+01

RESULT	ID	US-08-154-365-2	STANDARD:	PRT:	451 AA.
CC	CC	Sequence 2, Application US/08154365			
CC	CC	Patent No. 5624504			
CC	CC	GENERAL INFORMATION:			
CC	CC	APPLICANT: Dunne, Dana W.			
CC	CC	APPLICANT: Resnick, David			
CC	CC	APPLICANT: Kreiger, Monty			
CC	CC	APPLICANT: Joiner, Keith A.			
CC	CC	TITLE OF INVENTION: Method for Treating Gram-Positive			
CC	CC	NUMBER OF SEQUENCES: 2			
CC	CC	CORRESPONDENCE ADDRESS:			
CC	CC	ADDRESS: Patrea L. Pabst			
CC	CC	STREET: 1100 Peachtree Street, Suite 2800			
CC	CC	CITY: Atlanta			
CC	CC	STATE: Ga			
CC	CC	COUNTRY: USA			
CC	CC	ZIP: 30309-4530			
CC	CC	COMPUTER READABLE FORM:			
CC	CC	MEDIUM TYPE: floppy disk			
CC	CC	COMPUTER: IBM PC compatible			
CC	CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	CC	SOFTWARE: Patent Release #1.0, Version #1.25			
CC	CC	CURRENT APPLICATION DATA:			
CC	CC	APPLICATION NUMBER: US/08/154,365			
CC	CC	FILING DATE:			
CC	CC	CLASSIFICATION: 514			
CC	CC	ATTORNEY/AGENT INFORMATION:			
CC	CC	NAME: Pabst, Patrea L.			
CC	CC	REGISTRATION NUMBER: 31,284			
CC	CC	REFERENCE/DOCKET NUMBER: MIT6392			
CC	CC	TELECOMMUNICATION INFORMATION:			
CC	CC	TELEPHONE: (404)-815-6558			
CC	CC	TELEFAX: (404)-815-6555			
CC	CC	INFORMATION FOR SEQ ID NO: 2:			
CC	CC	SEQUENCE CHARACTERISTICS:			
CC	CC	LENGTH: 451 amino acids			
CC	CC	TYPE: amino acid			
CC	CC	STRANDEDNESS: single			

CC MOLECULE TYPE: protein
SQ SEQUENCE 585 AA; 65330 MW; 1874129 CN;
Query Match 46.1%; Score 363; DB 2; Length 585;
Best Local Similarity 43.0%; Pred. No. 2,08e-27;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;
Db 24 MRLADGATNCGRAVEIEYRGQWGVCDNLMDLTDASVVCRAFGFENATQALGRAAFQGS 83
Y 334 IRLAGKSGSHGRLVYRGQWGVCDGDTLNTYVVCRLQGFYKQQA-SANHFEST 392
Db 84 GPIWLDEVQCTGTASLADCKSLGMLKSNCRHERDAGVC 123
Y 393 GPIWLDEVSCGKETRFLQCSRRQWRHDCSHREDVSIAC 432
T 9
US-08-477-674-10 STANDARD: PRT: 585 AA.
xxxxxx
AC xxxxxx
XX
DT
XX
DE Sequence 10, Application US/08477674
XX
CC Sequence 10, Application US/08477674
CC Patent No. 5644035
CC GENERAL INFORMATION:
CC APPLICANT: Kochs, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casipit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-third street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/477,674
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE:
CC APPLICATION NUMBER: US/07/961,404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA; 65330 MW; 1874129 CN;

Query Match 46.1%; Score 363; DB 1; Length 585;
Best Local Similarity 43.0%; Pred. No. 2,08e-27;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;
Db 24 MRLADGATNCGRAVEIEYRGQWGVCDNLMDLTDASVVCRAFGFENATQALGRAAFQGS 83
Y 334 IRLAGKSGSHGRLVYRGQWGVCDGDTLNTYVVCRLQGFYKQQA-SANHFEST 392
Db 84 GPIWLDEVQCTGTASLADCKSLGMLKSNCRHERDAGVC 123
Y 393 GPIWLDEVSCGKETRFLQCSRRQWRHDCSHREDVSIAC 432
RESULT 10
ID US-08-794-795-7 STANDARD: PRT: 489 AA.
XX
XX
AC xxxxxx
XX
DT
XX
DE Sequence 7, Application US/08794795
XX
CC Sequence 7, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T.
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 489 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 489 AA; 49441 MW; 1141296 CN;
Query Match 44.7%; Score 352; DB 2; Length 489;
Best Local Similarity 42.4%; Pred. No. 2,99e-26;
Matches 42; Conservative 23; Mismatches 30; Indels 4; Gaps 3;
Db 394 VRIMGNT-NR-GRAEYVYNNEMGTICDDDDNNDATVFCMLGLSGNRALSS--YGGGSG 449

DE Sequence 2, Application US/08794795
XX
CC Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA; 49764 MW; 1218428 CN;
CC
CC Query Match
CC Best Local Similarity 44.1%; Score 347; DB 2; Length 495;
CC Matches 44; Conservative 21; Mismatches 30; Indels 4; Gaps 3;
DB 399 VRIVGS-SNR-GRAEYVYSTGTICDDDEWNSDAIVFCRMIG--YSKGRALYKVGAGTG 454
QY 334 IRLAGKGSHEGRLEYYRQWGTVCDDGWTETNTYVVCRLGFKYKQASANHFEESTG 393
DB 455 QIWLNVQCRGTESTLMSCTKNSGHHDCSHEDAGVEC 493
QY 394 PIVLDVSCSGKRETRFLQCSRQWRGHRHDCSHREDVSIAC 432
RESULT 14
ID US-08-794-795-6 STANDARD; PRT: 520 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08794795
XX
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:

CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA; 52658 MW; 1340662 CN;
CC
CC Query Match
CC Best Local Similarity 44.1%; Score 347; DB 2; Length 520;
CC Matches 44; Conservative 21; Mismatches 30; Indels 4; Gaps 3;
DB 424 VRIVGS-SNR-GRAEYVYSTGTICDDDEWNSDAIVFCRMIG--YSKGRALYKVGAGTG 479
QY 334 IRLAGKGSHEGRLEYYRQWGTVCDDGWTETNTYVVCRLGFKYKQASANHFEESTG 393
DB 480 QIWLNVQCRGTESTLMSCTKNSGHHDCSHEDAGVEC 518
QY 394 PIVLDVSCSGKRETRFLQCSRQWRGHRHDCSHREDVSIAC 432
RESULT 15
ID US-08-681-151-1 STANDARD; PRT: 356 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 1, Application US/08681151
XX
CC Sequence 1, Application US/08681151
CC Patent No. 5869637
CC GENERAL INFORMATION:
CC APPLICANT: Au-Young, Janice
CC APPLICANT: Bandman, Olga
CC APPLICANT: Braxton, Scott Michael
CC APPLICANT: Goll, Surya
CC TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

    hpp protein - protein database search, using Smith-Waterman algorithm
Run on:      Mon Mar 13 10:31:24 2000;      MasPar time 10.32 Seconds
Tabular output not generated.              457.141 Million cell updates/sec

```

```

Title: >US-09-147-947-6
Description: (334-433) from US09147947A.ppt (5 of 6)
Perfect Score: 987
Sequence: 1 IRLGGGSGHEGLEYVRS.....RQQRHDCSHREDVSTACY 100

```

Scoring table: PAM 150

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62

Statistics: Mean 37.820; Variance 61.366; scale 0.616

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	DB	ID	Description	Pred. No.
1	691	87.8	761	2	JC5759	brain-specific serine
2	419	53.2	918	2	JC4361	scavenger receptor Cy
3	413	52.3	453	2	S08276	LDL receptor I, macro
4	412	52.4	1116	2	S36077	M130 antigen - human
5	412	52.4	1149	2	I38006	M130 antigen (extracel
6	412	52.4	1151	2	I38004	M130 antigen (cytosol
7	412	52.4	1156	2	I38005	M130 antigen (cytosol
8	412	52.4	2153	2	T14893	M130 antigen (cytosol
9	407	51.7	451	2	A38415	scavenger receptor cy
10	405	51.5	454	2	I46862	macrophage scavenger
11	397	50.4	454	2	A44047	macrophage scavenger
12	393	49.9	458	2	B44407	macrophage scavenger
13	391	49.7	1436	2	A44966	antigen Wc1.1 precurs
14	374	47.5	600	2	S56742	mucin (clone pGM-1)
15	363	46.1	585	2	A47161	Mac-2-binding glycop
16	352	44.7	518	2	A35840	Macrophage bacteria-b
17	351	44.6	1290	2	A57190	ebherin precursor - r
18	350	44.5	577	2	A51202	Cytophilin C-associ
19	346	44.0	504	2	S56745	mucin (clone pGM1-1)
20	293	37.2	468	2	S26741	T-cell glycoprotein C
21	281	35.7	532	2	A33751	spectac receptor prec
22	281	35.7	626	2	I49100	mscd6 precursor - mu
23	114	14.5	495	1	S11270	T-cell surface glyco

45	106	13.5	1034	A53663	enterepoitidase (EC 3	9.57e-05
24	105	13.3	495	A26396	T-cell surface glycop	1.44e-04
25	106	13.5	495	A26396	gene Cb5 protein	1.61e-03
26	99	12.6	442	A47074	sh	1.61e-03
27	99	12.6	1019	A56318	enterepoitidase (EC 3	5.15e-03
28	96	12.2	1035	A43090	enterepoitidase (EC 3	5.25e-03
29	93	11.8	558	T15448	hypothetical protein	1.67e-02
30	92	11.7	202	T01605	blue copper-binding P	2.45e-02
31	91	11.6	613	S15468	complement C3b/C4b in	3.58e-02
32	91	11.6	2616	A57095	nudel protein precurs	3.58e-02
33	90	11.4	501	S4356	T-cell surface glycop	5.21e-02
34	89	11.3	2206	UC5280	voltage-dependent cal	7.57e-02
35	88	11.2	333	T15257	hypothetical protein	1.10e-01
36	88	11.2	494	A29079	lymphocyte surface gl	1.10e-01
37	88	11.2	1113	TJEO13	low-density lipoprote	1.10e-01
38	86	10.9	583	A29154	complement factor I (2.29e-01
39	84	10.7	2504	A57188	fatty-acid synthase (4.72e-01
40	84	10.7	2509	G01880	fatty-acid synthase (4.72e-01
41	83	10.5	411	G69534	probable acyl-CoA deh	6.75e-01
42	83	10.5	670	G71251	mannose-specific lect	9.61e-01
43	82	10.4	149	A58081	mannose-specific lect	9.61e-01
44	82	10.4	306	S50689	hypothetical protein	9.61e-01
45	82	10.4	340	C71266	conserved hypthetica	9.61e-01

ALIGNMENTS

[illegible]

```

oy      394 PIMLDVSCSGKETRFLOC SRQMGWRHDCSHREDVSIACY 433

RESULT      2
ENTRY      JC4361
TITLE      #type complete
           scavenger receptor Cys-rich epidermal growth factor precursor
           - sea lamprey
ORGANISM    #formal_name Petromyzon marinus #common_name sea lamprey
DATE        08-Jul-1996 #sequence_revision 08-Feb-1996 #text_change
           16-Jul-1999
ACCESSIONS  JC4361
REFERENCE   JC4361
#authors    Mayer, W.E.; Tichy, H.
#journal     Gene (1995) 164:267-271
#title       A cDNA clone from the sea lamprey Petromyzon marinus coding
           for a scavenger receptor Cys-rich (SRK) domain protein.
#cross-references NID:96069593
#accession  JC4361
#molecule-type mRNA
#residues   1-918 #label MAY
#cross-references GB:020652; NID:g790233; PID:g790234
           This protein is rich in cysteine and plays a role in intercellular
           contacts and cell activation or differentiation in the cytosome
           immune system.
CLASSIFICATION #superfamily EGF homology; scavenger receptor cysteine-rich
           domain homology
           glycoprotein; growth factor; receptor; transmembrane protein
KEYWORDS     #domain signal sequence #status predicted #label SIG\
FEATURE      #product scavenger receptor Cys-rich epidermal growth
           factor #status predicted #label MAY\
           #domain scavenger receptor cysteine-rich domain homology
           #label SRC1\
           #domain EGF homology #label EG1\
           #domain EGF homology #label EG2\
           #domain EGF homology #label EG3\
           #domain EGF homology #label EG4\
           #domain EGF homology #label EG5\
           #domain scavenger receptor cysteine-rich domain homology
           #label SRC2\
           #domain transmembrane #status predicted #label TM\
           #domain intracellular #status predicted #label INT\
           #binding_site carbohydrate (Asn) (covalent) #status
           predicted
SUMMARY      #length 918 #molecular-weight 101417 #checksum 2936

Query Match      53.2% Score 419; DB 2; Length 918;
Best Local Similarity 52.0% Pred. NO.2,42e-75;
Matches 52; Conservative 21; Mismatches 25; Indels 2; Gaps 2;

466 VALVG-GWCGQGEVEYYVYGSMGTVCDDSDMDRODAEYVQRLSCGYAVSAPNAFGVGS 524
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
0734 IRLAGKSGHEGLLEVYYGOMGTVCDDGTELTNTYYVVCQLGKRYKQASAN-HFEEST 392

Db 535 GOIWLDDVNCYGOESSLARCNSHSGWRHDCSHREDVSIAC 564
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 393 GPIWLDDVSCSGKETRFLOC SRQMGWRHDCSHREDVSIAC 432

RESULT      3
ENTRY      S08276
TITLE      #type complete
           LDL receptor I, macrophage - bovine
           macrophage scavenger receptor type I
           #formal_name Bos primigenius taurus #common_name cattle
           30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
           20-Sep-1999
ACCESSIONS  S08276
REFERENCE   S08276
#authors    Kodama, T.; Freeman, M.; Rohrer, L.; Zabrecky, J.;
           Matsuda, P.; Krieger, M.
           Nature (1990) 343:531-535
#journal     Nature I macrophage scavenger receptor contains alpha-helical
#title

```

[illegible]

921-1024	#domain scavenger receptor cysteine-rich domain homolog	#label SRC9	#molecular-weight 120979	#checksum 5687
SUMMARY				
Query Match	52.4%; Score 412; DB 2; Length 1116;			
Best Local Similarity 48.5%; Pred. No.1,24e-73;				
Matches 49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;				
Db	714 LRLVNGGRCAGREYIIEHSGSKITCDSDS-DLSDAHVCROLCGGEAINTGSAHEGEG	772		
QY	334 IRLAGGKSGHGRELVYRGMGTVCDDGWTGL-NRYVVCROLDGFKRYKGSAN-HEDES	391		
Db	773 TGPIMLDMKCNKESRIWCHSHGNGQONCRKEDAGYIC	813		
QY	392 TGPIMLDDVSCSGKETRFLOCSSRRQWRHDCSHREDVSIAC	432		
ENTRY	5			
TITLE	I38006 #type complete			
ORGANISM	M130 antigen (extracellular variant) - human			
DATE	17-May-1996 #sequence_rev1510N 17-May-1996 #text_change 02-Aug-1996			
ACCESSIONS	I38006; S35768			
REFERENCE	I38003			
#authors	Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.			
#journal	Eur. J. Immunol. (1993) 23:2320-2325			
#title	A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.			
#cross-references	MUTID:93380506			
#accession	I38006			
#status	Preliminary; translated from GB/EMBL/DBJ			
#molecule_type	mRNA			
#residues	1-1149 #label RES			
#cross-references	EMBL:Z29971; NID:g312147; PID:g312148			
CLASSIFICATION	#superfamily scavenger receptor cysteine-rich domain homolog			
FEATURE				
43-147	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC1\			
151-254	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC2\			
258-361	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC3\			
365-468	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC4\			
470-573	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC5\			
58-711	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC6\			
744-847	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC7\			
849-953	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC8\			
954-1057	#domain scavenger receptor cysteine-rich domain homolog			
	#label SRC9			
SUMMARY	#length 1149 #molecular-weight 124328 #checksum 487			
Query Match	52.4%; Score 412; DB 2; Length 1149;			
Best Local Similarity 48.5%; Pred. No.1,24e-73;				
Matches 49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;				
Db	747 LRLVNGGRCAGREYIIEHSGSKITCDSDS-DLSDAHVCROLCGGEAINTGSAHEGEG	805		
QY	334 IRLAGGKSGHGRELVYRGMGTVCDDGWTGL-NRYVVCROLDGFKRYKGSAN-HEDES	391		
Db	806 TGPIMLDMKCNKESRIWCHSHGNGQONCRKEDAGYIC	846		
QY	392 TGPIMLDDVSCSGKETRFLOCSSRRQWRHDCSHREDVSIAC	432		
RESULT	6			
ENTRY	I38004			
	#type complete			

TITLE	ENTRY	DATE	ACCESSIONS	REFERENCE	KEYWORDS	FEATURE
M130 antigen (cytosolic variant 1) - human						
#formal_name Homo sapiens #common_name man						
17-May-1996 #sequence_revision 17-May-1996 #text_change						
07-Feb-1997						
I38004; S36078						
I38003						
Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;						
Willits, A.C.; Mason, D.Y.						
Eur. J. Immunol. (1993) 23:2320-2325						
A new macrophage differentiation antigen which is a member of						
the scavenger receptor superfamily.						
#cross-references MVID:93380506						
#accession I38004						
#status preliminary; translated from GB/EMBL/DBJ						
#molecule_type mRNA						
#residues 1-1151 #label RES						
#cross-references EMBL:Z22969; NID:g312144; PID:g312144						
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology						
KEYWORDS cytosol						
43-147						
151-254						
258-361						
365-468						
470-573						
575-678						
711-814						
816-920						
921-1024						
SUMMARY						
#length 1151 #molecular_weight 124820 #checksum 1481						
Query Match 52.4%; Score 412; DB 2; Length 1151;						
Best Local Similarity 48.5%; Pred. NO.1,24e-73;						
Matches 49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;						
Db 714 LRLVNGGCGRAGREYIEHESGWTICDDSDM-DLSDAHVYRQLGCGEALNATGSAHFGEG 772						
Y 334 IRLAGGKSGHEGLLEYRYRGQNTVCDDEWTEL-NTIYVCRQLGKRYGKQASAN-HFEES 391						
Db 773 TGPIMLDEMCKNGKESRIWCHSGHGQGNCRHKEDAGYIC 813						
Y 392 TGPIMLDVSCSGKETRFLOCSRNRQGRHDCSHREDVSIAC 432						
RESULT 7						
ENTRY I38005 #type complete						
TITLE M130 antigen (cytosolic variant 2) - human						
ORGANISM #formal_name Homo sapiens #common_name man						
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change						
07-Feb-1997						
I38005; S36079						
I38003						
ACCESSIONS						
REFERENCE						
#authors						
#journal						
#title						
#cross-references MVID:93380506						
#accession I38005						
#status preliminary; translated from GB/EMBL/DBJ						
#molecule_type mRNA						
#residues 1-1156 #label RES						
#cross-references EMBL:Z22970; NID:g312145; PID:g312146						
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology						

RESULT	9	
ENTRY	A38415	#type complete
TITLE	macrophage scavenger receptor splice form I - human	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change	
	20-Sep-1999	
ACCESSIONS	A38415; A44408	
REFERENCE	Matsumoto, A.; Naito, M.; Itakura, H.; Ikemoto, S.; Asaoka, H.; Hayakawa, I.; Kanamori, H.; Aburatani, H.; Takaku, F.; Suzuki, H.; Kobari, Y.; Miyai, T.; Takahashi, K.; Cohen, E.H.; Wydro, R.; Housman, D.E.; Kodama, T. Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9133-9137	
#journal	Human macrophage scavenger receptors: primary structure, expression, and localization in atherosclerotic lesions.	
#title	MUSD.91067661	
#cross-references	MIMD:91067661	
#accession	A38415	
#status	Preliminary	
##molecule_type	mRNA	
##residues	1-451 #label MAT	
##cross-references	GB:D90187; NID:g219989; PIDN:BA14208.1; PTD:01014913; PID:g219990	
REFERENCE	A44408	
#authors	Emi, M.; Asaoka, H.; Matsumoto, A.; Itakura, H.; Kurihara, Y.; Wade, Y.; Kanamori, H.; Yazaki, Y.; Takahashi, E.; Lepert, M.; Lalouel, J.; Kodama, T.; Mukai, T. J. Biol. Chem. (1993) 268:2120-2125	
#journal	Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene.	
#title	macrophage scavenger receptor gene.	
#cross-references	MIMD:93131971	
#accession	A44408	
#status	Preliminary: not compared with conceptual translation	
##molecule_type	nucleic acid	
##residues	155-272 #label EMI	
##note	sequence extracted from NCBI backbone (NCBI:P:123189)	
GENETICS		
#gene	GDB:MSR1	
##cross-references	GDB:128046; OMIM:153622	
#map_position	8p22-8p22	
CLASSIFICATION	#superfamily unassigned collagens; scavenger receptor cysteine-rich domain homology	
KEYWORDS	alternative splicing; coiled coil; transmembrane protein	
FEATURE		
347-450	#domain scavenger receptor cysteine-rich domain homology	
SUMMARY	#length 451 #molecular_weight 49762 #checksum 6168	
Query Match	51.7%; Score 407; DB 2; Length 451;	
Best Local Similarity	51.5%; Pred. No. 2, 0.7e-72;	
Matches	52; Conservative 18; Mismatches 28; Indels 3; Gaps 2;	
Db	350 VRLVGGSGPHBGRVEILHSQSGTICDDREKRVAGVYCRSLGYP-GVQAYIKAAHFEQSG 408	
QY	334 IRLAGGSGSHBGRLEFYRRGQMGTCVDDGTELTNTVYRCQDGFYKQQA--SANHFEE 391	
Db	409 TGPITLNEVCFPGHSEIECKIRONGTRACSHSEAGATC 449	
QY	392 TGPITLDDVSCSGRETRFLQCSRRQGRHDCSHREDVSTAC 432	
RESULT	10	
ENTRY	I46862	#type complete
TITLE	macrophage scavenger receptor type I - rabbit	
ORGANISM	#formal_name Oryctolagus cuniculus #common_name domestic rabbit	
DATE	14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change	
	20-Sep-1999	
ACCESSIONS	I46862	
REFERENCE	Bickel, P.E.; Freeman, M.W. J. Clin. Invest. (1992) 90:1450-1457	
#authors	Rabbit aortic smooth muscle cells express inducible	
#title	Rabbit aortic smooth muscle cells express inducible	

```

TITLE      macrophage scavenger receptor, MSRI - mouse
ALTERNATE_NAMES
ORGANISM   #formal_name Mus musculus #common_name mouse
DATE       30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
          20-Sep-1999
ACCESSIONS B44407; A38260; I56334
REFERENCE  #authors
          #journal
          #title
          #cross-references PMID:93131972
          #accession B44407
          ##status preliminary; not compared with conceptual translation
          ##molecule_type nucleic acid
          ##residues 1-458 ##label DOI
          ##experimental_source macrophage-like cell line P38D
          ##note sequence extracted from NCBI backbone (NCBIP:123207)
          #authors
          #journal
          #title
          #cross-references PMID:91062370
          #accession A38260
          ##status preliminary
          ##molecule_type mRNA
          ##residues 349-458 ##label FRE
          ##cross-references GB:M59445; GB:M36817; NID:g192737; PIDN:AAA37464.1;
          PID:g192738
REFERENCE  I56334
          #authors Ashkenas, J.; Penman, M.; Vasile, E.; Acton, S.; Freeman,
          M.W.; Krieger, M.
          #journal J. Lipid Res. (1993) 34:983-1000
          #title Structures and high and low affinity ligand binding
          properties of murine type I and type II macrophage
          scavenger receptors.
          #cross-references PMID:93359822
          #accession I56334
          ##status preliminary; translated from GB/EMBL/DBJ
          ##molecule_type mRNA
          ##residues 5-458 ##label RES
          ##cross-references GB:L04574; NID:g293745; PIDN:AAA39747.1; PID:g293746
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor
          cysteine-rich domain homology
FEATURE
          354-457 #domain scavenger receptor cysteine-rich domain homology
SUMMARY
          #length 458 #molecular_weight 50130 #checksum 1435
          Query Match 49.9%; Score 393; DB 2; Length 458;
          Best Local Similarity 47.0%; Pred. No. 5,31e-69;
          Matches 47; Conservative 18; Mismatches 34; Indels 1; Gaps 1;
          Db 357 VRLVGSGAHEGRVETIHOGQWGTICDDRWDIRAGOVYCRSLGYOEVLAVHRAHFGOGT 416
          :||| |||:|||||:::|||||:||||| |||| |||: |||:|
          Qy 334 IRLAGKGSHEBLLEVYRGQWGTVCDDGTELTNYVVCRLQGFKKYQQA-SANFEEST 392
          :|||:|::|::|::| ||| ||| |||:|::|:|
          Db 417 GPIPLNEVMCFGRSSSIENCKINQKQVLSCSISDEAGVTC 456
          :|||:|::|::|::| ||| ||| |||:|::|:|
          Qy 393 GPIPLDVSCKGKTRFRLQCSRQWRHDCSHREDVSIAC 432
          :|||:|::|::|::| ||| ||| |||:|::|:|
RESULT      13
ENTRY
          A46496 #type complete
          antigen WCL.1 Precursor - bovine
          #organism Bos primigenius taurus #common_name cattle
          #formal_name Bos primigenius taurus #common_name mouse
          18-Jun-1993 #sequence_revision 19-May-1994 #text_change

```


##cross-references GB:X79089; NID:9483473; PIDN:CA55699.1; PID:9483474

REFERENCE

#authors S32384

#journal Iacobelli, S.; Buccì, I.; d'Egidio, M.; Giuliani, C.; Natoli, C.; Tinari, N.; Rubinstein, M.; Schlessinger, J.

#title FEBS Lett. (1993) 319:59-65

#cross-references MUID:93202277

#accession S32384

##molecule_type protein

##residues 19-24,'C',26-40 ##label IN2

REFERENCE PC2211

#authors Inohara, H.; Raz, A.

#journal Biochem. Biophys. Res. Commun. (1994) 201:1366-1375

#title Identification of human melanoma cellular and secreted ligands for galectin-3.

#cross-references MUID:94296411

#accession PC2211

##molecule_type protein

##residues 19-38 ##label INO

#note the reported molecular weight of this protein was 98K

#accession PC2212

##molecule_type protein

##residues 19-26 ##label IN2

#note the reported molecular weight of this protein was 70K

REFERENCE A41005

#authors Rosenbery, I.; Cherayil, B.J.; Isselbacher, K.J.; Pillai, S.

#journal J. Biol. Chem. (1991) 266:18731-18736

#title Mac-2-binding glycoproteins. Putative ligands for a cytosolic beta-galactoside lectin.

#cross-references MUID:92011634

#accession A41005

##molecule_type protein

##residues 19-22,'G',24-27,'XX',30 ##label ROS

#note the reported molecular weight of this protein, designated Mac-2-binding glycoprotein 2 (M2BP-2), was 70K; the amino-terminal sequence of a related form of 98K, designated M2BP-1, could not be determined

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology

KEYWORDS extracellular protein; glycoprotein

FEATURE 1-18

21-124

#domain signal sequence #status predicted #label SIG

#domain scavenger receptor cysteine-rich domain homology

#label SRC

SUMMARY #length 585 #molecular-weight 65330 #checksum 2263

IV Match 46.1%; Score 363; DB 2; Length 585;

t Local Similarity 43.0%; Pred.No 9,75e-62;

Ghes 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;

Db 24 MRLADGATNGRVEIIFRGQGTVCNLDLTPDASVYCRALGFENATQALGRAAFGGS 83

QY 334 IRLAGGKSHHGRLVYRGQGTVCDDGWELMTYVYCRQLGFKYGRQA-SANHFEST 392

Db 84 GPIMLDEVQCTGTASLADCKSLGLKSNCRHERDAGVVC 123

QY 393 GPIMLDVSCSGKETRFQCSRRQWGRHDCSHREDVSIAC 432

Search completed: Mon Mar 13 10:31:36 2000

Job time : 12 secs.

THIS PAGE BLANK (USPTO)

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:30:22 2000; Maspar time 6.44 Seconds
 463,783 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-147-947-6
 Description: (334-433) from US09147947A.pap (5 of 6)
 Perfect Score: 787
 Sequence: 1 IRLAGKSGHGRLEVVYRG.....RQGRHDCSHREDVSIACY 100

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 38.688; Variance 57.261; scale 0.676

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
787	100.0	875	1	NETR_HUMAN	1.39e-183
691	87.8	761	1	NEUROTRYP SIN PRECURSOR	7.85e-157
413	52.5	433	1	MSRE_BOVIN	7.34e-81
407	51.7	451	1	MACROPHAGE SCAVENGER R	2.93e-79
405	51.5	454	1	MSRE_RABIT	1.00e-78
393	49.9	458	1	MACROPHAGE SCAVENGER R	1.56e-75
391	49.7	1436	1	MC1L_BOVIN	5.32e-75
293	37.2	468	1	CD6_HUMAN	2.22e-49
281	35.7	532	1	SPER_STRPU	2.60e-46
114	14.5	495	1	CD5_BOVIN	4.72e-07
106	13.5	1034	1	ENTR_PIG	1.75e-05
105	13.3	495	1	CD5_HUMAN	2.72e-05
100	12.7	492	1	TMS2_HUMAN	2.40e-04
100	12.7	1069	1	ENTR_MOUSE	2.40e-04
99	12.6	1019	1	ENTR_HUMAN	3.68e-04
96	12.2	1035	1	ENTR_BOVIN	1.31e-03
91	11.6	2616	1	NDL_DROME	1.04e-02
90	11.4	491	1	CD5_RAT	1.55e-02
88	11.2	494	1	CD5_MOUSE	3.46e-02
86	10.9	583	1	CEAL_HUMAN	7.62e-02
84	10.7	2504	1	FAS_HUMAN	1.66e-01
82	10.4	306	1	YE16_YEAST	3.55e-01
80	10.2	756	1	RIR1_HABIN	7.51e-01

RESULT	1	ALIGNMENTS
ID	NETR_HUMAN	STANDARD; PRT; 875 AA.
AC	P56730;	
DT	15-DEC-1999 (Rel. 39, Created)	
DT	15-DEC-1999 (Rel. 39, Last sequence update)	
DT	15-DEC-1999 (Rel. 39, Last annotation update)	
DE	NEUROTRYP SIN PRECURSOR (EC 3.4.21.-) (MOTOP SIN).	
GN	PRSS12.	
OS	Homo sapiens (Human)	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;	
NC	Eutheria; Primates; Catarrhini; Homidae; Homo.	
NC	[1]	
RP	SEQUENCE FROM N.A.	
RP	TISSUE-BRAIN;	
RX	MEDLINE: 96201705.	
RA	PROBA K., GSCHWEND T.P., SONDERGGER P.;	
RT	"Cloning and sequencing of the cDNA encoding human neurotysin."	
RL	Biochim. Biophys. Acta 1396:143-147 (1998).	
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).	
CC	- SUBCELLULAR LOCATION: SECRETED.	
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.	
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.	
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).	
CC	EMBL: AJ001531; CA04816.1; -	
DR	PROSITE: PS00134; TRYPSIN_HIS. 1.	
DR	PROSITE: PS00135; TRYPSIN_SER. 1.	
DR	PROSITE: PS00420; SERPACT_RECEPTOR. 3.	
KW	Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.	
FT	SIGNAL	1 20
FT	CHAIN	21 875
FT	DOMAIN	23 92
FT	DOMAIN	93 161
FT	DOMAIN	170 271
FT	DOMAIN	280 381
FT	DOMAIN	387 487
FT	DOMAIN	387 487

334-433 / SCARCS
 Use any 8 bits No.3

```
FT DOMAIN 500 601 SRCR 4.
FT DOMAIN 619 875 SERINE PROTEASE.
FT ACT_SITE 619 830 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA; 97011 MW; 67D52272B CRC32;

Query Match 100.0%; Score 787; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.39e-183;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 387 IRLAGKSGHSEGRLEYYRGQMGTVCDGTELTNTYVVCROLGFKYKQASANHFESTG 446
    IRLAGKSGHSEGRLEYYRGQMGTVCDGTELTNTYVVCROLGFKYKQASANHFESTG 393
334 IRLAGKSGHSEGRLEYYRGQMGTVCDGTELTNTYVVCROLGFKYKQASANHFESTG 393
447 PIMDDVSCSGKETRFLOCSSRQWGRHDCSHREDVSIACY 486
    PIMDDVSCSGKETRFLOCSSRQWGRHDCSHREDVSIACY 433
394 PIMDDVSCSGKETRFLOCSSRQWGRHDCSHREDVSIACY 433

RESULT 2 STANDARD; PRT; 761 AA.
ID NETR_MOUSE
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEOTRYPSEIN PRECURSOR (EC 3.4.21.-) (MOTOPSLIN) (BRAIN-SPECIFIC SERINE
DE PROTEASE 3) (BSSP-3).
GN PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthera; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GSCHWEND T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neotrypsin, a novel multidomain serine protease expressed in the
RT nervous system."
RL Mol. Cell. Neurosci. 9:207-219(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSUBOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
RT single-like structure and three scavenger receptor cysteine-rich
RT motifs."
RL Biochem. Biophys. Res. Commun. 239:386-392(1997).
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=LUNG;
RX MEDLINE: 90136965.
RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
RA KRIGER M.;
RT "Type I macrophage scavenger receptor contains alpha-helical and
RT collagen-like coiled coils."
RL Nature 343:531-535(1990).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=LUNG;
RX MEDLINE: 90136973.
RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIGER M.;
RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
RT scavenger receptor type II."
RL Nature 343:570-572(1990).
CC -!- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROSCLEROSIS
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE
CC ENOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
```

```
DR EMBL: Y13192; CAAT3646.1; -
DR EMBL: D89871; BAA23986.1; -
DR MGD; MG1:1100881; PRSS12.
DR PFAM; PF00530; SRCR; 3.
DR PFAM; PF00089; trypsin; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00420; SPPRAC1_RECEPTOR; 3.
KW Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 761
FT DOMAIN 85 157 NEOTRYPSEIN.
FT DOMAIN 166 267 KRINGLE.
FT DOMAIN 273 373 SRCR 1.
FT DOMAIN 386 487 SRCR 2.
FT DOMAIN 505 761 SRCR 3.
FT DOMAIN 505 516 SERINE PROTEASE.
FT ACT_SITE 516 517 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 562 562 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 CHARGE RELAY SYSTEM.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA; 84118 MW; 3F3C4F35 CRC32;

Query Match 87.8%; Score 691; DB 1; Length 761;
Best Local Similarity 82.0%; Pred. No. 7.85e-157;
Matches 82; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 273 IRLAGKSGHSEGRLEYYRGQMGTVCDGTELTNTYVACRLGFRYKQSSVNHFDGSR 332
    IRLAGKSGHSEGRLEYYRGQMGTVCDGTELTNTYVACRLGFRYKQSSVNHFDGSR 393
334 IRLAGKSGHSEGRLEYYRGQMGTVCDGTELTNTYVACRLGFRYKQSSVNHFDGSR 393
333 PIMDDVSCSGKEVFQICSRQWGRHDCSHREDVGLICY 372
    PIMDDVSCSGKEVFQICSRQWGRHDCSHREDVGLICY 433
394 PIMDDVSCSGKETRFLOCSSRQWGRHDCSHREDVSIACY 433

RESULT 3 STANDARD; PRT; 453 AA.
ID MSRE_BOVIN
AC P21758;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
DE LDL RECEPTOR I AND II).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=LUNG;
RX MEDLINE: 90136965.
RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
RA KRIGER M.;
RT "Type I macrophage scavenger receptor contains alpha-helical and
RT collagen-like coiled coils."
RL Nature 343:531-535(1990).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=LUNG;
RX MEDLINE: 90136973.
RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIGER M.;
RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
RT scavenger receptor type II."
RL Nature 343:570-572(1990).
CC -!- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROSCLEROSIS
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE
CC ENOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
```


RP SEQUENCE FROM N.A.
RX MEDLINE: 91067661.
RA MATSUMOTO A., NAITO M., ITAKURA H., IKEMOTO S., ASAKOHA H.,
RA HAKAKAWA I., KANMORI H., AUBRAVANI H., TAKAKU F., SUZUKI H.,
RA KOHARI Y., MIYAI T., TAKAHASHI C., COHEN E. H., WIDRO R.

DT 01-oct-1996 (Rel. 34, Created)
DT 01-oct-1996 (Rel. 34, last sequence update)
DT 01-oct-1996 (Rel. 34, last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
LDL RECEPTOR I AND II)

```

GN MSRI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Theria; Lagomorpha; Leporidae; Oryctolagus.
RN RP
   (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 93016877.
RA BICKEL P.E., FREEMAN M.W.;
RT "Rabbit aortic smooth muscle cells express inducible macrophage
RT scavenger receptor messenger RNA that is absent from endothelial
RL J. Clin. Invest. 90:1450-1457(1992).
   [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93131972.
RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
RA MIYAZAKI T., NAKAMURA H., UESUGI S., IMANISHI T., KAWABE Y.,
RA ITAKURA H., YAZAKI Y., MATSUMOTO A.;
RT "Charged collagen structure mediates the recognition of negatively
RT charged macromolecules by macrophage scavenger receptors.";
   J. Biol. Chem. 268:2126-2133(1993).
CC CC
   - I - FUNCTION: MEMBRANE GLYCOPROTEIN'S IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
CC CC
   - I - SUBUNIT: HOMOTRIMER.
CC CC
   - I - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC CC
   - I - ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
CC OF THE SAME GENE.
-----
CC CC
   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to licens@sdb-sdb.ch).
-----
CC CC
   EMBL; L11693; AAA31402.1; -
DR EMBL; L11692; AAA31403.1; -
DR EMBL; D13381; BAA02649.1; -
DR PROSITE; PS00420; SPBRCT_RECEPTOR_1.
DR PFAM; PF00530; SRCR_1.
DR PFAM; PF01391; Collagen_1.
KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
KV Heptad repeat pattern; Receptor; Alternative splicing.
FT DOMAIN 1 50
FT TRANSMEM 51 73
   (POTENTIAL).
   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
   (POTENTIAL).
   EXTRACELLULAR (POTENTIAL).
   SPACER (PROBABLE).
   ALPHA-HELICAL COILED-COIL.
   COLLAGEN-LIKE.
   CYS-RICH.
FT DOMAIN 74 109
FT DOMAIN 110 272
FT DOMAIN 273 344
FT DOMAIN 345 454
FT CARBOHYD 82 82
FT CARBOHYD 102 102
FT CARBOHYD 143 143
FT CARBOHYD 184 184
FT CARBOHYD 221 221
FT CARBOHYD 249 249
FT CARBOHYD 267 267
FT POTENTIAL.
FT POTENTIAL.
FT POTENTIAL.
FT TPSTAVR -> RPYOLTPT (IN ISOFORM II).
FT VARSPLIC 355 454
FT MISSING (IN ISOFORM II).
FT CONFLICT 106 106
   H -> D (IN REF. 2).
SO SEQUENCE 454 AA; 49745 MW; 5D780348 CRC32;

```

	Y	392	TGPIMLDVVSCSGKRETRFLQCSRROMGRHDCSHREDVSIAAC	432
	D	412	TGPIMLNEVPCLTGMSSIECKIRPMGVAVCVSHGEDAGVC	452
	O	392	TGPIMLDVVSCSGKRETRFLQCSRROMGRHDCSHREDVSIAAC	432
RESULT	6			
ID	MSRE_MOUSE	STANDARD:	PRT:	458 AA.
AC	P30204;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	MAROPHAEE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED LLDI RECEPTOR I AND II).			
DE	MSRI OR SCVR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 93359822.			
RA	ASHKENAS J., PENMAN M., VASILJE E., ACTON S., FREEMAN M.W., KRUEGER M.:			
RT	"Structures and high and low affinity ligand binding properties of murine type I and type II macrophage scavenger receptors."			
RL	J. Lipid Res. 34:983-1000(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A. (SHORT FORM).			
RX	MEDLINE: 93131972.			
RA	DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y., ITAKURA H., YATAKI Y., MATSUMOTO A.:			
RA	MAYAKA H., YATAKI Y., MATSUMOTO A.:			
RT	"Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors."			
RL	J. Biol. Chem. 268:2126-2133(1993).			
RN	[3]			
RP	SEQUENCE OF 349-458 FROM N.A.			
RX	MEDLINE: 91062370.			
RA	FREEMAN M., ASHKENAS J., REES D.J., KINGSLLEY D.M., COPELAND N.G., RA JENKINS N.A., KRUEGER M.:			
RT	"An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors."			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).			
RN	[4]			
RP	SEQUENCE OF 1-4 FROM N.A.			
RX	MEDLINE: 95395388.			
RA	AFFRING R.P., FREEMAN M.W.:			
RT	"Structure of the murine macrophage scavenger receptor gene and evaluation of sequences that regulate expression in the macrophage cell line, P388D."			
RL	J. Lipid Res. 36:1305-1314(1995).			
-I-	FUNCTION: MEMBRANE GLYCOPROTEIN IMPLICATED IN THE PATHOLOGIC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.			
CC	TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING MODIFIED LOW DENSITY LIPOPROTEINS (LDL).			
CC	-I- SUBUNIT. HOMOTRIMER.			
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.			
CC	-I- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MAROPHAEE SCAVENGER RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.			
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).			

SO SEQUENCE 468 AA: 49828 MW: 042D5AC8 CRC32:

Query Match 37.2%; Score 293; DB 1; Length 468;

Best Local Similarity 44.1%; Pred. No. 2,22e-49;
Matches 45: Conservative 25; Mismatches 26; Indels 6; Gaps 6;

DB 161 LRLVGGGACAGVEMLEHEKMSVCDTWT-DLEDHAYVCRQLGCGNVAALPGLHPTPG 219

QY 334 IRLAGKSGHEGLLEYYRGQWGTVCDDGTELT-NTYVCRQLGFKYKQA-SANHEES 391

DB 220 RPIHRDVCNCGAEE-YLMDCEGLP-GOHYCGKEHDAGVVC 259

QY 392 TGPIMLDVSCSGKETRFL-QCSRWGRHDCSHREDVSIAC 432

RESULT 9
ID SPER STRPU STANDARD: PRT: 532 AA.

AC P16264:

01-AUG-1990 (rel. 15, Created)

01-AUG-1990 (rel. 15, Last sequence update)

05-JUL-1999 (rel. 38, Last annotation update)

EGG PEPTIDE SPECTR RECEPTOR PRECURSOR.

Strongylocentrotus purpuratus (Purple sea urchin).

Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;

Euechinozoa; Echinacea; Echinoida; Strongylocentrotidae;

Strongylocentrotus.

[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 477-489.

RX MEDLINE: 89184581.

RA DANOTT L.J., JORDAN J.E., BELET R.A., GARBERS D.L.;

"Cloning of the mRNA for the protein that crosslinks to the egg

peptide spectr."

Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).

CC -1- FUNCTION: RECEPTOR FOR THE EGG PEPTIDE SPECTR.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL: J04518; AAA30078.1; -

DR PIR: A32751; A32751.

DR PROSITE: PS00420: SPECTR_RECEPTOR. 4.

DR PFAM: PF00530; SRCR; 4.

KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.

CHAIN 1 30

DOMAIN 31 532

TRANSMEM 492 520

DOMAIN 521 532

DOMAIN 40 488

REPEAT 147 147

REPEAT 148 260

REPEAT 375 375

REPEAT 376 488

CARBOHYD 78 78

CARBOHYD 115 115

CARBOHYD 459 459

SEQUENCE 532 AA: 57820 MW: D27174AE CRC32:

Query Match 35.7%; Score 281; DB 1; Length 532;

Best Local Similarity 43.7%; Pred. No. 2.60e-46;

Matches 45: Conservative 19; Mismatches 34; Indels 5; Gaps 5;

DB 264 IRLMDSGPHEGVETIWDHDMGTICDDGMDMADANVVCROAGYKAGVSKGKDEDFGF 323

QY 334 IRLAGKSGHEGLLEYYRGQWGTVCDDGTELTNTYVCRQLGFKYKQA-SANHEE-S- 391

DB 324 IYAPHTSFVMTGVEDRLIDCILRDGMT-HSCYHVEDASVVC 365

QY 392 T-GPIMLDVSCSGKETRFLQCSRRQ-WGRHDCSHREDVSIAC 432

RESULT 10

ID CD5 BOVIN STANDARD: PRT: 495 AA.

AC P19238;

DT 01-NOV-1990 (rel. 16, Created)

DT 01-NOV-1990 (rel. 16, Last sequence update)

DT 01-NOV-1997 (rel. 35, Last annotation update)

DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCYTE GLYCOPROTEIN

CD5, T1/LEU-1) (LYMPHOCYTE ANTIGEN CD5).

GN CD5.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

OC Bovinae; Bos.

[1]

SEQUENCE FROM N.A.

TISSUE-THYMUS;

RA YU Q., REICHERT M., BROUSSEAU T., CLEUTER Y., BURRY A., KETTMANN R.;

"Sequence of bovine CD5."

Nucleic Acids Res. 18:5296-5296(1990).

CC -1- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL

PROLIFERATION. CD5 INTERACTS WITH CD72/LYB-2.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- SIMILARITY: CONTAINS 2 SRCR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL: X53061; CAA37231.1; -

DR PIR: S11270; S11270.

DR PFAM: PF00530; SRCR; 3.

KW Signal; Transmembrane; Glycoprotein; T-cell.

CHAIN 1 24

DOMAIN 25 495

TRANSMEM 373 402

DOMAIN 403 495

DOMAIN 33 134

DOMAIN 273 382

DISULFID 43 108

CARBOHYD 51 51

CARBOHYD 117 117

CARBOHYD 323 323

CARBOHYD 369 369

SEQUENCE 495 AA: 54333 MW: F8ED2AE0 CRC32:

Query Match 14.5%; Score 114; DB 1; Length 495;

Best Local Similarity 37.2%; Pred. No. 4.72e-07;

Matches 16: Conservative 12; Mismatches 14; Indels 1; Gaps 1;

DB 34 MRLSGSGRCQGRLEYSNGTWTAVHSQSGQLSYQVAPROF 76

QY 334 IRLAGKSGHEGLLEYYRGQWGTVCDDGTELTNTYVCRQL 375

RESULT 11

ID ENTK_PIG STANDARD: PRT: 1034 AA.

AC P98074;

DT 01-FEB-1996 (rel. 33, Created)

DT 01-FEB-1996 (rel. 33, Last sequence update)

DT 15-JUL-1998 (rel. 36, Last annotation update)

DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).

GN PRSS7 OR ENTK.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Catartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-DUODENAL MUCOSA;
 RX MEDLINE: 94327548.
 RA MATSUOKA M., ICHINOSE M., YAHAGI N., KAKEI N., TSUKADA S.,
 MIKI K., KUROKAWA K., TASHIRO K., SHIOKAWA K., SHINOMIYA K.,
 UMEYAMA H., INOUE H., TAKAHASHI T., TAKAHASHI K.;
 RL "Structural characterization of porcine enteropeptidase";
 J Biol Chem. 269:19976-19982(1994).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 TRYPSINOGEN
 CC -1- SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
 (HEAVY) CHAIN, AND A MINI CHAIN.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 NAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D30799; BAA06459.1; -.
 CC HSSP: P00763; LDPO.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 PFAM: PF00057; Ldl_recept_a; 2.
 PFAM: PF00089; trypsin; 1.
 PFAM: PF00431; CUB; 2.
 PFAM: PF00530; SRCR; 1.
 PFAM: PF00629; MAM; 1.
 PFAM: PF01380; SEA; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 KW Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 52 117
 FT CHAIN 118 799
 FT CHAIN 800 1034
 FT TRANSMEM 19 47
 FT DOMAIN 197 238
 FT DOMAIN 240 349
 FT DOMAIN 357 519
 FT DOMAIN 539 649
 FT DOMAIN 656 694
 FT DOMAIN 693 786
 FT ACT_SITE 840 840
 FT ACT_SITE 891 891
 FT ACT_SITE 986 986
 FT LIPID 2 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236

FT DISULFID 658 670 BY SIMILARITY.
 FT DISULFID 665 683 BY SIMILARITY.
 FT DISULFID 677 692 BY SIMILARITY.
 FT DISULFID 787 911 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 825 841 BY SIMILARITY.
 FT DISULFID 925 992 BY SIMILARITY.
 FT DISULFID 956 971 BY SIMILARITY.
 FT DISULFID 982 1010 BY SIMILARITY.
 FT CARBOHYD 116 116 POTENTIAL.
 FT CARBOHYD 147 147 POTENTIAL.
 FT CARBOHYD 170 170 POTENTIAL.
 FT CARBOHYD 194 194 POTENTIAL.
 FT CARBOHYD 283 283 POTENTIAL.
 FT CARBOHYD 343 343 POTENTIAL.
 FT CARBOHYD 350 350 POTENTIAL.
 FT CARBOHYD 403 403 POTENTIAL.
 FT CARBOHYD 455 455 POTENTIAL.
 FT CARBOHYD 485 485 POTENTIAL.
 FT CARBOHYD 518 518 POTENTIAL.
 FT CARBOHYD 549 549 POTENTIAL.
 FT CARBOHYD 645 645 POTENTIAL.
 FT CARBOHYD 697 697 POTENTIAL.
 FT CARBOHYD 701 701 POTENTIAL.
 FT CARBOHYD 721 721 POTENTIAL.
 FT CARBOHYD 740 740 POTENTIAL.
 FT CARBOHYD 761 761 POTENTIAL.
 FT CARBOHYD 804 804 POTENTIAL.
 FT CARBOHYD 863 863 POTENTIAL.
 FT CARBOHYD 902 902 POTENTIAL.
 FT CARBOHYD 964 964 POTENTIAL.
 SO SEQUENCE 1034 AA; 114776 MW; 24386471 CRC32;
 Query Match 13.5%; Score 106; DB 1; Length 1034;
 Best Local Similarity 27.4%; Pred. No. 1.75e-05;
 Matches 17; Conservative 17; Mismatches 27; Indels 1; Gaps 1;
 Db 693 VREFLNGTANSGVDFRISVHTACAEWNTTQSDVCCOLLGIGTGN-SMPEFFSSGG 751
 QY 334 IRLAGGKSHGHEVRYRQGWTVCDGWTDLNTYVCHQDLGKFKYKQASNNFEESTG 393
 Db 752 PF 753
 QY 394 PI 395
 RESULT 12
 ID CD5_HUMAN STANDARD; PRT; 495 AA.
 AC P06127;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCYTE GLYCOPROTEIN
 DE TI/LEU-1) (LYMPHOCYTE ANTIGEN CD5).
 GN CD5 OR LEU1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87014786.
 RA JONES N.H., CLABY M.L., DIALYNAS D.P., HUAG H.-J.S.,
 RA HERZENBERG L.A., STROWINGER J.L.;
 RT "Isolation of complementary DNA clones encoding the human lymphocyte
 RT glycoprotein TI/Leu-1".
 RL Nature 323:346-349(1986).
 RN [2]
 RP INTERACTION WITH CD72/LYB-2.
 RX MEDLINE: 91270374.
 RA VAN DE VELDE H., VON HOEGEN I., LEO W., PARNES J.R., THIELEMANNS K.;
 RT "The B-cell surface protein CD72/Lyb-2 is the ligand for CD5.";
 RL Nature 351:662-665(1991).
 CC -1- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL
 CC PROLIFERATION. CD5 INTERACTS WITH CD72/LYB-2.

```

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 SRCR DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD5 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd5.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04391; CAA27979.1; .
CC DR PIR: A26396; A26396.
CC DR MIM: 153340; .
CC DR PFM: PF00530; SRCR: 3.
CC KW Signal; Transmembrane; Glycoprotein; T-cell.
CC -----
CC SIGNAL 1 24
CC TRANSMEM 25 495 T-CELL SURFACE GLYCOPROTEIN CD5.
CC DOMAIN 25 372 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 373 402 POTENTIAL.
CC DOMAIN 403 495 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 34 133 SRCR 1.
CC FT DOMAIN 275 382 SRCR 2.
CC FT DISULFID 44 107 POTENTIAL.
CC FT CARBOHYD 116 116 POTENTIAL.
CC FT CARBOHYD 241 241 POTENTIAL.
CC SQ SEQUENCE 495 AA; 54625 MW; 8507C34C CRC32;

Query Match 13.3%; Score 105; DB 1; Length 495;
Best Local Similarity 37.9%; Pred. No. 2,72e-05;
Matches 11; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

Db 36 RLRRSNKCGOGLVYIKDGMHWCSQSW 64
QY 335 RLGGKSGHGRLEVYRGQWGTVCDDGW 363

RESULT 13
ID TMS2_HUMAN STANDARD; PRT; 492 AA.
AC 015393;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TRANSMEMBRANE PROTEIN, SERINE 2 (EC 3.4.21.-).
GN TMS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC [1]
OC SEQUENCE FROM N.A.
OC MEDLINE: 97468144.
OC PROONT-GINGCOPINO A., CHEN H., PEITSCH M.C., ROSSIER C.,
OC ANTONMARKIS S.E.,
OC "Cloning of the TMS2 gene, which encodes a novel serine protease
OC with transmembrane, IDLRA, and SRCR domains and maps to 21q22.3."
OC Genomics 44:309-320(1997).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND
CC WEALY IN SEVERAL OTHER TISSUES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
CC EMBL: U75329; AAC51784.1; .
CC DR HSSP: P00763; IDPO.
CC DR MIM: 602060; .
CC DR PROSITE: PS00134; TRYPSIN_HIS: 1.
CC DR PROSITE: PS00135; TRYPSIN_SER: 1.
CC DR PROSITE: PS01209; IDLRA_1; 1.
CC DR PROSITE: PS01068; IDLRA_2; 1.
CC DR PFM: PF00057; IdL-recept_a; 1.
CC DR PFM: PF00089; trypsin; 1.
CC KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
CC -----
CC DOMAIN 1 84
CC TRANSMEM 85 105 POTENTIAL.
CC FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 111 150 LDL-RECEPTOR CLASS A.
CC FT DOMAIN 151 243 SRCR.
CC FT DOMAIN 256 492 CATALYTIC.
CC FT ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 113 126 BY SIMILARITY.
CC FT DISULFID 120 139 BY SIMILARITY.
CC FT DISULFID 133 148 BY SIMILARITY.
CC FT DISULFID 244 365 BY SIMILARITY.
CC FT DISULFID 281 297 BY SIMILARITY.
CC FT DISULFID 410 426 BY SIMILARITY.
CC FT DISULFID 437 465 BY SIMILARITY.
CC FT CARBOHYD 213 213 POTENTIAL.
CC FT CARBOHYD 249 249 POTENTIAL.
CC SQ SEQUENCE 492 AA; 53847 MW; 19BC1F67 CRC32;

Query Match 12.7%; Score 100; DB 1; Length 492;
Best Local Similarity 40.7%; Pred. No. 2,40e-04;
Matches 11; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Db 165 RKSNHPYCDQDMNENGRACRDMGK 191
QY 352 RGQWGTVCDDGWTENLYVVCROLGFK 378

RESULT 14
ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC [1]
OC SEQUENCE FROM N.A.
OC STRAIN-C57BL/6; TISSUE-DUODENUM.
OC YUAN X., LU D., ROBIN D.C., PUNG C.Y.M., SADLER J.E.;
OC Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

```

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U73378; AAB37317.1; -

DR HSSP: P00763; IDPO

DR MGD: MG1197523; PRSS7.

DR PROSITE: PS00134; TRYPsin_HIS; 1.

DR PROSITE: PS00135; TRYPsin_SER; 1.

DR PROSITE: PS01180; CUB; 2.

DR PROSITE: PS00740; MAM_1; 1.

DR PROSITE: PS01209; LDLRA_1; 2.

DR PROSITE: PS00068; LDLRA_2; 2.

DR PFAM: PF00057; 1d_recept_a; 2.

DR PFAM: PF00089; trypsin; 1.

DR PFAM: PF00431; CUB; 2.

DR PFAM: PF00530; SRCR; 1.

DR PFAM: PF00629; MAM; 1.

DR PFAM: PF01390; SEA; 1.

DR Signal-anchor: Glycoprotein: Myristate; Hydroxylase;

DR Serine protease; Zymogen; Transmembrane; Repeat;

DR CHAIN 1 829

DR CHAIN 1 829

DR TRANSMEM 19 47

DR DOMAIN 227 268

DR DOMAIN 270 379

DR DOMAIN 387 549

DR DOMAIN 569 679

DR DOMAIN 686 724

DR DOMAIN 733 816

DR ACT_SITE 874 874

DR ACT_SITE 925 925

DR ACT_SITE 1021 1021

DR LIPID 2 2

DR DISULFID 229 242

DR DISULFID 236 255

DR DISULFID 249 266

DR DISULFID 688 700

DR DISULFID 695 713

DR DISULFID 707 722

DR DISULFID 817 945

DR DISULFID 859 875

DR DISULFID 959 1027

DR DISULFID 991 1006

DR DISULFID 1017 1045

DR CARBOHYD 147 147

DR CARBOHYD 197 197

DR CARBOHYD 212 212

DR CARBOHYD 373 373

DR CARBOHYD 380 380

DR CARBOHYD 433 433

DR CARBOHYD 515 515

DR CARBOHYD 579 579

DR CARBOHYD 675 675

DR CARBOHYD 727 727

DR CARBOHYD 751 751

DR CARBOHYD 770 770

DR CARBOHYD 791 791

DR CARBOHYD 897 897

DR CARBOHYD 936 936

DR CARBOHYD 999 999

DR SEQUENCE 1069 AA; 118735 MW; 488825A2 CRC32;

Query Match 12.7%; Score 100; DB 1; Length 1069;
 Best Local Similarity 24.2%; Pred. No. 2.40e-04;

Matches 23; Conservative 27; Mismatches 40; Indels 5; Gaps 4;
 Db 723 VELNLTGNSNMGVFNHISHIACAEWNTQISNEVCHLLGL--GSANSPISSTGG 780
 QY 334 IRLAGKSGSHERLEVYNGQVGTCDQDWTLENTYVQRLGFKGKASNHPESTG 393
 Db 781 GPFVRYNQAPNGSLITPSLQCSQDSLLILQCNH 815
 QY 394 -P-IWLDVSCSGKE-TRELCSSRRQWRHDCSHR 425

RESULT 15
 ID ENTK_HUMAN STANDARD; PRT; 1019 AA.
 AC P98073;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-DUODENUM;
 RX MEDLINE; 95234679.
 RA KITAMOTO Y., VEILE R.A., DONIS-KELLER H., SADLER J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 RT the proteolytic activator of trypsinogen.";
 RL Biochemistry 34:4562-4568(1995).
 RN [2]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE-DUODENUM;
 RX MEDLINE; 94329561.
 RA KITAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 RT protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
 CC TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
 CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
 CC -1- SIMILARITY: CONTAINS 2 CUB-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U09860; AAC50138.1; -
 DR HSSP: P00763; IDPO.
 DR MIM: 226200; -
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.

```
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PFAM; PF00057; ldl_recept_a; 2.
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00431; CUB; 2.
DR PFAM; PF00530; SRCR; 1.
DR PFAM; PF00629; MAM; 1.
DR PFAM; PF01390; SEA; 1.
KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;
  Serine protease; Zymogen; Transmembrane; Repeat;
  NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 1 784
FT CHAIN 19 47
FT TRANSMEM 182 223
FT DOMAIN 225 334
FT DOMAIN 342 504
FT DOMAIN 524 634
FT DOMAIN 641 679
FT ACT_SITE 678 771
FT ACT_SITE 825 825
FT ACT_SITE 876 876
FT ACT_SITE 971 971
FT LIPID 2 2
FT LIPID 184 197
FT DISULFID 191 210
FT DISULFID 204 221
FT DISULFID 643 655
FT DISULFID 650 668
FT DISULFID 662 677
FT DISULFID 772 896
FT DISULFID 810 826
FT DISULFID 910 977
FT DISULFID 941 956
FT DISULFID 967 995
FT CARBOHYD 116 116
FT CARBOHYD 147 147
FT CARBOHYD 179 179
FT CARBOHYD 328 328
FT CARBOHYD 335 335
FT CARBOHYD 388 388
FT CARBOHYD 440 440
FT CARBOHYD 470 470
FT CARBOHYD 503 503
FT CARBOHYD 534 534
FT CARBOHYD 630 630
FT CARBOHYD 682 682
FT CARBOHYD 706 706
FT CARBOHYD 725 725
FT CARBOHYD 848 848
FT CARBOHYD 887 887
FT CARBOHYD 909 909
FT CARBOHYD 949 949
SQ SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;

Query Match 12.6%; Score 99; DB 1; Length 1019;
Best Local Similarity 31.3%; Pred.No. 3.68e-04;
Matches 20; Conservative 14; Mismatches 25; Indels 5; Gaps 4;

Db 678 VREFNGTNNNG-L-VREFQSIWHTACAEWNTQTISNDVQCLGLGSGNSSKPI-FSTD 734
OY 334 IRLAGKSGSHGRLVYRGQ--WGTVCDGDWTELTNYVVCROLGFKYGRQASANHFEE 391

Db 735 GGPF 738
OY 392 TGPI 395
```

Search completed: Mon Mar 13 10:30:31 2000
Job time : 9 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

h_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:30:49 2000; Maspar time 16.65 seconds
Tabular output not generated. 416.315 Million cell updates/sec

```

```
Title: >US-09-147-947-6
Description: (334-433) from US09147947A.ppt (5 of 6)
Perfect Score: 787
Sequence: 1 IRLGAGGSHGRLFVYIRG.....RQMRHDCSHREDVSTACY 100
```

Scoring table: PAM 150

Searched: 225878 seqs, 69334122 residues

```
post-processing: Minimum Match 08
                  Listing first 45 summaries
```

```
Database:      spiremb112
              1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
              5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_oranella
              9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
              13:sp_vertebrate 14:sp_virus
```

Statistics: Mean 37.816; Variance 60.520; scale 0.625

SUMMARIES

Query No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	432	54.9	2043	5	096843	SRCR DOMAIN, MEMBRANE	1 878-787
2	419	53.2	1918	13	092098	PEMA SRCR PROTEIN PREC	3 176-767
3	412	52.4	1116	4	007898	M130 ANTIGEN PRECURSOR	1 728-747
4	412	52.4	1149	4	007901	M130 ANTIGEN, EXTRACEL	1 728-747
5	412	52.4	1151	4	007899	M130 ANTIGEN, CYTOSOL	1 728-747
6	412	52.4	1156	4	007900	M130 ANTIGEN, CYTOSOL	1 728-747
7	412	52.4	2153	5	097375	SCAVENGER RECEPTOR CYS	1 728-747
8	383	48.7	359	6	029110	SCAVENGER RECEPTOR PRO	2 468-67
9	383	48.7	774	4	094404	LYSLYL OXIDASE-RELATED	2 468-67
10	380	48.3	546	6	029111	SCAVENGER RECEPTOR PRO	1 344-66
11	374	47.5	600	6	028910	MUCIN (FRAGMENT) .	3 996-667
12	369	46.9	1534	6	095218	HENSIN.	6 696-667
13	365	46.4	462	6	028881	MEMBRANE PROTEIN SCAVE	6 388-667
14	364	46.3	578	11	P70117	PANCREAS CANCER-ASSOCI	1 128-656
15	363	46.1	585	4	008380	MAC-2 BINDING PROTEIN	1 978-656
16	362	46.0	895	13	094621	LYSLYL OXIDASE RELATED	3 445-667
17	362	46.0	1785	4	094409	DMRT1/SKB.1. PROTEIN PR	3 445-667
18	362	46.0	1785	4	094711	DMRT1 PROTEIN.	3 445-667
19	361	45.9	1036	11	070513	MAMA.	6 086-657
20	360	45.7	1574	5	097378	SCAVENGER RECEPTOR CYS	1 066-661

21	353	44.9	352.11	035301	SP-ALPHA.	5.42e-60
22	350	44.9	352.11	035300	SP-ALPHA.	5.42e-60
23	352	44.7	518.11	060754	BACTERIA BINDING MACRO	9.50e-60
24	351	44.6	1290.11	062827	EBERININ.	1.66e-59
25	350	44.5	577.11	007797	PEPTIDYLPROLYL ISOMERA	2.92e-59
26	350	44.5	2083.11	060997	CAR-DUCIN PRECURSOR	2.92e-59
27	347	44.1	520.4	095553	MACROPHAGE RECEPTOR.	1.57e-58
28	346	44.0	504.6	028808	MUCIN (FRAGMENT).	2.74e-58
29	343	43.6	347.4	034866	SP ALPHA.	1.47e-57
30	339	43.1	483.11	090WB9	MACROPHAGE RECEPTOR MA	1.37e-56
31	339	43.1	578.6	035649	CYCLOPHILIN C-ASSOCIAT	1.37e-56
32	332	42.2	480.6	029112	SCAVENGER-RECEPTOR PRO	6.81e-55
33	331	42.1	804.6	029113	SCAVENGER-RECEPTOR PRO	1.19e-54
34	326	41.4	356.6	092762	UNKNOWN MRNA, PARTIAL.	1.19e-54
35	326	41.4	754.11	092175	LYSYL OXIDASE-RELATED	1.92e-53
36	320	40.7	437.6	029109	SCAVENGER-RECEPTOR PRO	5.37e-52
37	318	40.4	127.6	095516	M130 ANTIGEN (FRAGMENT	1.63e-51
38	308	39.1	528.5	097379	SCAVENGER RECEPTOR CYS	4.10e-49
39	308	39.1	531.5	017064	SPB5.	4.10e-49
40	300	38.1	822.13	09YHC1	LYSYL OXIDASE HOMOLOG	3.16e-47
41	299	38.0	638.4	09Y5Y8	LYSYL OXIDASE-LIKE PRO	5.83e-47
42	293	37.2	592.4	09Y4K7	CD6E.	1.57e-45
43	293	37.2	595.4	09Y4K9	CD6C.	1.57e-45
44	293	37.2	601.4	09Y4K8	CD6D.	1.57e-45
45	293	37.2	635.4	09Y4L0	CD6B.	1.57e-45

ALIGNMENTS

ID	RESULT	1	PRELIMINARY;	PRT:	2043 AA.	DB
AC	096943;					
DT	01-MAY-1999 (TREMBlrel. 10, Created)					
DT	01-MAY-1999 (TREMBlrel. 10, Last sequence update)					
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)					
DE	SRCR DOMAIN, MEMBRANE FORM 2.					
GN	SRCRM2.					
OS	Geodia cydonium (Sponge).					
OC	Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;					
OC	Astrophoridae; Geodiidae; Geodia.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 98369060.					
RA	BLUMBACH B., PANCER Z., DIEHL-SEIFERT B., STEFFEN R., MUEKKNER J.,					
RA	MUELLER I., MUELLER W.E.G.;					
RT	"The putative sponge aggregation receptor. Isolation and					
RT	characterization of a molecule composed of scavenger receptor					
RT	cysteine-rich domains and short consensus repeats."					
RL	J. Cell Sci. 111:2635-2644 (1998).					
RL	EMBL; Y14953; CAA75175.1; "					
DR	HSSP; P10998; 1VVC.					
DR	PROSITE; PS00420; SPERACT_RECEPTOR; 8.					
SO	SEQUENCE 2043 AA; 220896 MW; 4CE19401 CRC32;					
Query Match 54.9%; Score 432; DB 5; Length 2043;						
Best Local Similarity 51.0%; Pred. No. 1,87e-79;						
Matches 51; Conservative 21; Mismatches 27; Indels 1; Gaps 1;						
Db	936	IRLGGSGPHEGRYEIYQCVGTCVDDSGCPDADAVYCRQLGYANASRATVBAEFGRT	995			
Qy	334	IRLAGGSGSHGRLEVEYRQGWGVDDGTELTNTYVVCNQLDFKRYGKQNSA-NHFEEST	392			
Db	996	GETIWDNVACTGFPENSIDECRSNGWDHNCGRHEDAGAVC	1035			
Qy	393	GPIWLDVSCSGKTRFLQCSRRQMGWRHDCSHREDVSIAIC	432			
RESULT	2	PRELIMINARY;	PRT;	918 AA.		
ID	092098;					
AC	092098;					
DT	01-NOV-1996 (TREMBlrel. 01, Created)					
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)					
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)					

Query	Subject	Score	DB	Length	Indels	Gaps
Db	714 LRLVNGGRCAGRAVEIYHEGSGMGTICDSDW-DLSDAHVVCRLQCGEAINATGSAHFGEG	52.4%	DB 4	1156	3	3
Qy	334 IRLAGKGSHEGRLEVLVYRGWGVTCDGWTETL-NTYVVCRLQCFKYGKQASAN-HFEES	48.5%	Pred. NO. 1.72e-74	25	3	3
Db	773 TGPWLDEMCKNGKRSRIIMOCSHWGQONCRHKEDAGVIC	52.4%	DB 4	1156	3	3
Qy	392 TGPWLDDVSCSGKETRFLQCSRRQWGRHDCSHREDVSIAC	45.2%	Pred. NO. 1.72e-74	25	3	3

```

DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.
GN SCRL12.
OC Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota: Metazoa: Echinodermata: Echinozoa: Echinoidea;
OC Euechinozoa: Echinozoa: Echinozoa: Echinozoa: Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RA PANCER Z., RAST J., DAVIDSON E.H.
RT "Evolution of the Immune System: Transcription Factors and Downstream
RT Genes of the Mammalian Immune System Expressed in Sea Urchin
RL Coelomocytes."
RL Immunogenetics 0:0-0(1999).
DR EMBL: AF064259; AAD08654.1.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 15.
KW Signal; Receptor.
FT SIGNAL 1 16
FT CHAIN 17 2153
FT FT
FT SEQUENCE 2153 AA; 226557 MW; CA3BA8A9 CRC32;
SQ
Query Match 52.4%; Score 412; DB 5; Length 2153;
Best Local Similarity 48.0%; Pred. No. 1,72e-74;
Matches 48; Conservative 25; Mismatches 26; Indels 1; Gaps 1;
Db 684 MRLVGGGNSRGRLEISINNGVTCDDSDINDATVVCROLGSSAVSAPTSANHGGS 743
Qy 334 IRLAGKGSGHEGRLEVVYRGQMGVCDGDTLNTVYVCHQLGKRYK-KQASANHFEE 392
Db 744 GTIMLDVYSCAGNENSLMDGCHRLGRLVHNCANAHADAIVYC 783
Qy 393 GPWLDVYSCSGKETRFLOCSRRQMGHRDCHREDEVSIAC 432
RESULT 8
ID 029110 PRELIMINARY; PRT; 369 AA.
AC 029110;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SCAVENGER-RECEPTOR PROTEIN (FRAGMENT).
GN WCI.
OS Sus scrofa (pig).
OC Eukaryota: Metazoa: Chordata: Cranialia: Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-SLAB/B;
RX MEDLINE; 97364683.
RA KANAN J.H.C., NAYEM N., BINNS R.M., CHAIN B.M.;
RT "Mechanisms for variability in a member of the scavenger-receptor
RT Cysteine-rich superfamily."
RL Immunogenetics 46:276-282(1997).
DR EMBL: X99333; CA67707.1;
DR PROSITE: PS00420; SPERACT_RECEPTOR; 2.
DR PFAM: PF00530; SRCR; 3.
FT NON_TER 1
FT NON_TER 369
FT SEQUENCE 369 AA; 39710 MW; C665BBA4 CRC32;
SQ
Query Match 48.7%; Score 383; DB 6; Length 369;
Best Local Similarity 48.0%; Pred. No. 2,46e-67;
Matches 49; Conservative 23; Mismatches 25; Indels 5; Gaps 5;
Db 35 LRLVDG-GSYCSGVEILHOGSGWTCVDDSM-DIDAHVVCROLDGCGKITISALGSAHFCA 92
Qy 334 IRLAGKGSGH-EGRLVEVYRGQMGVCDGDTLNTVYVCHQLGKRYK-KQASANHFEE 390
Db 93 GSGRIWLDVYSCGKESHLMQCPRGQGNCHHKEDAGVIC 134
Qy 391 STGPWLDVYSCSGKETRFLOCSRRQMGHRDCHREDEVSIAC 432

```

RESULT 9
ID 09Y4K0 PRELIMINARY: PRT: 774 AA.
AC 09Y4K0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE LYSYL OXIDASE-RELATED PROTEIN.
GN WS9-14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA SATO H., PAPACONSTANTINOU J., SATO H., GOLDSTEIN S.;
RT "Regulation of a Novel Gene Encoding a Lysyl Oxidase-Related Protein
in Cellular Adhesion and Senescence."
RL J. Biol. Chem. 0:0-0(1997).
RL EMBL: U89942; AAB49697.1; -
RL PROSITE: PS00420; SPBRCT_RECEPTOR.1.
SQ SEQUENCE 774 AA; 86724 MW; 1A5C47AA CRC32;

Query Match 48.7%; Score 383; DB 4; Length 774;
Best Local Similarity 47.5%; Pred. No. 2,46e-67;
Matches 48; Conservative 22; Mismatches 29; Indels 2; Gaps 2;
DB 58 IRLAGKRRHSGREVEYVDGQGVCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
QY 334 IRLAGKRGSH-EGRLVYRGQGVCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
DB 118 EGPIMLDNHTGNTGNTLAAGTNGVGTCDCKTEPDVGYVC 158
QY 392 GPTWLDVSCSGKTRFLQCSRRQGRHDCSHREDVSIAC 432

RESULT 10
ID 029111 PRELIMINARY: PRT: 546 AA.
AC 029111;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE SCAVENGER-RECEPTOR PROTEIN PRECURSOR (FRAGMENT).
GN WCL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-SLAB/B;
RC MEDLINE: 97364683.
RA KANAN J.H.C., NATEEM N., BINNS R.M., CHAIN B.M.;
RT "Mechanisms for variability in a member of the scavenger-receptor
cysteine-rich superfamily."
RL Immunogenetics 46:276-282(1997).
DR EMBL: X99334; CA67708.1; -
DR PFAM: PF00530; SRCR; 5.
KW Signal.
FT SIGNAL 1 22
FT CHAIN 23 >546
FT NON_TER 546
SQ SEQUENCE 546 AA; 59112 MW; F98A0F1F.CRC32;

Query Match 48.3%; Score 380; DB 6; Length 546;
Best Local Similarity 46.5%; Pred. No. 1.34e-66;
Matches 47; Conservative 25; Mismatches 26; Indels 3; Gaps 3;

DB 371 IRLVGESECSRVVELLHGGSGTICDESN-DIDDAHYVCRQLGCGDAISALGSAPFGVG 429
QY 334 IRLAGKRGSH-EGRLVYRGQGVCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
DB 430 SGRVLDVNCGTGKSHLWOCPEFRGGRHDCRHKEDAGVYC 470
QY 392 GPTWLDVSCSGKTRFLQCSRRQGRHDCSHREDVSIAC 432

RESULT 11
ID 028910 PRELIMINARY: PRT: 600 AA.
AC 028910;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MUCIN (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 95374471.
RA NUNES D.P., KEATES A.C., AFDPAL N.H., OFFNER G.D.;
RT "Bovine gall-bladder mucin contains two distinct tandem repeating
sequences: evidence for scavenger receptor cysteine-rich repeats."
RL Biochem. J. 310:41-48(1995).
RL EMBL: S78981; AAB35069.1; -
DR PFAM: PF00530; SRCR; 4.
FT NON_TER 1
SQ SEQUENCE 600 AA; 63999 MW; 32CE8265.CRC32;

Query Match 47.5%; Score 374; DB 6; Length 600;
Best Local Similarity 50.5%; Pred. No. 3.99e-65;
Matches 51; Conservative 19; Mismatches 28; Indels 3; Gaps 3;
DB 256 IRLVNSDRCQGVVELYRGSGVGTCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
QY 334 IRLAGKRGSH-EGRLVYRGQGVCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
DB 316 GPTWLDVSCSGKTRFLQCSRRQGRHDCSHREDVSIAC 432
QY 393 GPTWLDVSCSGKTRFLQCSRRQGRHDCSHREDVSIAC 432

RESULT 12
ID 095218 PRELIMINARY: PRT: 1594 AA.
AC 095218;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HENSIN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 97096804.
RA TAKITO J., HIKITA C., AL-AWOATI O.;
RT "Hensin, a new collecting duct protein involved in the in vitro
RT plasticity of intercalated cell polarity."
RL J. Clin. Invest. 98:2324-2331(1996).
DR EMBL: AF043112; AAD02242.1; -
DR HSSP: P29392; ISFP.
DR PROSITE: PS00420; SPBRCT_RECEPTOR.1.
DR PFAM: PF00530; SRCR; 2
SQ SEQUENCE 1594 AA; 172763 MW; 34EB5962.CRC32;

Query Match 46.9%; Score 369; DB 6; Length 1594;
Best Local Similarity 46.5%; Pred. No. 6.69e-64;
Matches 47; Conservative 22; Mismatches 29; Indels 3; Gaps 3;

DB 821 IRLVNGSTRQGVVELYRGSGVGTCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
QY 334 IRLAGKRGSH-EGRLVYRGQGVCDDEFSIAAHVCRGLGYEAKSWTSSYKGG 117
DB 881 GSFVLDVSCSGQEP-YLMNCGRHRLSHNGHYDAVYC 920
QY 393 GPTWLDVSCSGKTRFLQCSRRQGRHDCSHREDVSIAC 432

```
RESULT 13
ID Q2881 PRELIMINARY; PRT; 462 AA.
AC Q2881;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE MEMBRANE PROTEIN SCAVENGER RECEPTOR HOMOLOG (FRAGMENT).
GN T19.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
[1]
SEQUENCE FROM N.A.
MEDLINE: 95169648.
O'KEEFE M.A., METCALFE S.A., GLEN M.D., BONDEN T., MCINNES S.,
KIMPTON W.G., CAHILL R.N., HEIN W.R., WALKER I.D.,
"lymph node homing cells biologically enriched for gamma delta T cells
express multiple genes from the T19 repertoire."
RT Int. Immunol. 6:1687-1697(1994).
RL EMBL: S76313; AAB33543.1; -.
DR EMBL: S76311; AAB33543.1; JOINED.
DR PROSITE; PS00420; SPERACT_RECEPTOR; 1.
DR PFAM; PF00530; SRCR; 3.
FT NON_TER 1
FT SEQUENCE 462 AA; 49776 MW; 63EAEDB2 CRC32;

Query Match 46.4%; Score 365; DB 6; Length 462;
Best Local Similarity 43.6%; Pred. No. 6,38e-63;
Matches 44; Conservative 26; Mismatches 28; Indels 3; Gaps 3;

Db 4 LRLVGGCPCAGREILIDOXSMXICDDGW-DLDDAYVVCROGCEPALNATGSAHAGAG 62
QY 334 IRLAGGSGSHGRLEVYRGOMGTVCDDGTELN-TTVVCRQLGFKYG-KQASNHFEES 391
OY 392 GPIWLDVSCSGKETRFLOCSRQWRGRHDCSHREDVSIAC 432

RESULT 14
ID P70117 PRELIMINARY; PRT; 578 AA.
AC P70117;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PANCREAS CANCER-ASSOCIATED PROTEIN 4.
Mesorhizocetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae; Mesocricetus.
RN [1]
SEQUENCE FROM N.A.
RA SCHAEFFERT C., POUR P.M., MACDONALD R.G., CHANEY W.G.,
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U73375; AAB18745.1; -.
DR PROSITE; PS00420; SPERACT_RECEPTOR; 1.
DR PFAM; PF00530; SRCR; 1.
DR PRINTS; PR00258; SPERACTRCPTR.
SQ SEQUENCE 578 AA; 64398 MW; 002E8BD2 CRC32;

Query Match 46.3%; Score 364; DB 11; Length 578;
Best Local Similarity 43.0%; Pred. No. 1.12e-62;
Matches 43; Conservative 22; Mismatches 34; Indels 1; Gaps 1;

Db 24 MRLVNGASANEGRVEIFYRGOMGTVCNDLMDLTADSVYCRALGFENATQALGRAAFGGR 83
QY 334 IRLAGGSGSHGRLEVYRGOMGTVCDDGTELN-TTVVCRQLGFKYGKQKQ-SANHFEEST 392
OY 392 GPIWLDVSCSGKETRFLOCSRQWRGRHDCSHREDVSIAC 432

Db 84 GPVVLDEVECTGTEPSLANCSLSGLKSCRGCEKDAVVC 123
QY 393 GPIWLDVSCSGKETRFLOCSRQWRGRHDCSHREDVSIAC 432
```

```
RESULT 15
ID Q08380 PRELIMINARY; PRT; 585 AA.
AC Q08380;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE MAC-2 BINDING PROTEIN PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE: 93300818.
RX KOTHS K., TAYLOR E., HALENEBECK R., CASIPIT C., WANG A.;
RT "Cloning and characterization of a human Mac-2-binding protein, a new
RT member of the superfamily defined by the macrophage scavenger receptor
RT cysteine-rich domain."
RL J. Biol. Chem. 268:14245-14249(1993).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE: 94308070.
RA ULLRICH A., SURES I., D'EGIDO M., JALLAL B., POWELL T.J., HERBST R.,
RA DREPS A., AZAM M., ROBINSTEIN M., NATOLI C.;
RT "The secreted tumor-associated antigen 90K is a potent immune
RT stimulator."
RL J. Biol. Chem. 269:18401-18407(1994).
DR EMBL: L13210; AAA6193.1; -.
DR EMBL: X79089; CAA5699.1; -.
DR PFAM; PF00530; SRCR; 1.
DR PRINTS; PR00258; SPERACTRCPTR.
KW Signal.
FT SIGNAL 1
FT CHAIN 19 POTENTIAL.
FT SEQUENCE 585 AA; 65330 MW; AAER9E32 CRC32;

Query Match 46.1%; Score 363; DB 4; Length 585;
Best Local Similarity 43.0%; Pred. No. 1.97e-62;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;

Db 24 MRLDGGAINQGRVEIFYRGOMGTVCNDLMDLTADSVYCRALGFENATQALGRAAFGGR 83
QY 334 IRLAGGSGSHGRLEVYRGOMGTVCDDGTELN-TTVVCRQLGFKYGKQKQ-SANHFEEST 392
OY 393 GPIWLDVSCSGKETRFLOCSRQWRGRHDCSHREDVSIAC 432

Db 84 GPVVLDEVECTGTEPSLANCSLSGLKSCRGCEKDAVVC 123
QY 393 GPIWLDVSCSGKETRFLOCSRQWRGRHDCSHREDVSIAC 432
```

Search completed: Mon Mar 13 10:31:07 2000
Job time : 18 secs.

THIS PAGE BLANK (USPTO)

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA: 65330 MW: 1874129 CN;
SQ
Query Match 51.7%; Score 391; DB 2; Length 585;
Best Local Similarity 52.0%; Pred. No. 6,80e-30;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;
Db 24 MRLADGATNCRGVEIFYGQMGCTVCDNLMDLTDASVYCRALGFENATQALGRAAFGGG 83
OY 447 VRLMDGKNKKEGVVEFYINGQWGTICDDGWTDKDAVICRQLGYGPARARMAFYEGGK 506
Db 84 GPTLMDVQCTGTASLADCKSLGWLKSNCRHEDAGVVC 123
OY 507 GPIHVDNVKCTGNERSLADCIKQDIGHNCRHSEDAGVIC 546
XX
XX US-08-477-674-10 STANDARD; PRT: 585 AA.
XX
XX xxxxxx
DE Sequence 10, Application US/08477674
CC Patent No. 5644035
CC GENERAL INFORMATION:
CC APPLICANT: Koths, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casipit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/477,674
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC FILING DATE: 15-OCT-1992
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA: 65330 MW: 1874129 CN;
SQ
Query Match 51.7%; Score 391; DB 1; Length 585;
Best Local Similarity 52.0%; Pred. No. 6,80e-30;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;
Db 24 MRLADGATNCRGVEIFYGQMGCTVCDNLMDLTDASVYCRALGFENATQALGRAAFGGG 83
OY 447 VRLMDGKNKKEGVVEFYINGQWGTICDDGWTDKDAVICRQLGYGPARARMAFYEGGK 506
Db 84 GPTLMDVQCTGTASLADCKSLGWLKSNCRHEDAGVVC 123
OY 507 GPIHVDNVKCTGNERSLADCIKQDIGHNCRHSEDAGVIC 546
XX
XX RESULT 3
XX ID US-08-473-791-10 STANDARD; PRT: 585 AA.
XX
XX xxxxxx
DE Sequence 10, Application US/08473791
CC Patent No. 5735340
CC GENERAL INFORMATION:
CC APPLICANT: Koths, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casipit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/473,791
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC FILING DATE: 15-OCT-1992
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein


```

CC      TITLE OF INVENTION:  eptor
CC      NUMBER OF SEQUENCES:  9
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE:  SmithKline Beecham Corporation
CC      STREET:  709 Swedeland Road
CC      CITY:  King of Prussia
CC      STATE:  PA
CC      COUNTRY:  USA
CC      ZIP:  19406
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE:  Diskette
CC      COMPUTER:  IBM Compatible
CC      OPERATING SYSTEM:  DOS
CC      SOFTWARE:  FASTSEQ for Windows Version 2.0
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER:  US/08/794,795
CC      FILING DATE:  04-FEB-1997
CC      CLASSIFICATION:  435
CC      PRIORITY APPLICATION DATA:
CC      APPLICATION NUMBER:  ATG50009P
CC      FILING DATE:  22-MAY-1996
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME:  Han, William T
CC      REGISTRATION NUMBER:  34,344
CC      REFERENCE/DOCKET NUMBER:  ATG50009
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE:  610-270-5219
CC      TELEFAX:  610-270-4026
CC      TELEX:
CC      INFORMATION FOR SEQ. ID NO:  7:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH:  489 amino acids
CC      TYPE:  amino acid
CC      STRANDEDNESS:  single
CC      TOPOLOGY:  linear
CC      MOLECULE TYPE:  protein
CC      SEQUENCE  489 AA;  49441 MW;  1141296 CN;
SQ
Query Match  48.7%; Score 369; DB 2; Length 489;
Best Local Similarity 54.0%; Pred. No. 1,31e-27;
Matches  54; Conservative  17; Mismatches  24; Indels  5; Gaps  2.

Db  394 VRINGGNNR--GRAEYVYNNMEWGTCDDDDNNNDATYFCMLGY---SRGALSTYGGGS 448
QY  447 VRLDGEKKKRGREVEINQGWGTCDDGWTDKDAAYICRQLGKGPARTAYAFEEGK 506
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db  449 GNIMLDNVNCGTENSIMDCSKNSGWHNCHVHNEDAGVEG 488
QY  507 GPIHDVNYKTGNERSLADCLIKODIGRHNRHSHSDAGVIC 546
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT  6
ID  US-08-392-367B-2      STANDARD;          PRT:  518 PA.
AC      xxxxxx
DT
DT
XX      Sequence 2, Application US/08392367B
DE
XX      Sequence 2, Application US/08392367B
XX      Patent No. 5691197
CC      GENERAL INFORMATION:
CC      APPLICANT:  Trygsvason, Karl
CC      APPLICANT:  Elomaa, Outi
CC      APPLICANT:  Kangas, Maarit
CC      TITLE OF INVENTION:  An Insolated DNA Sequence For a
CC      Patent No. 5691197
CC      TITLE OF INVENTION:  No. 5691197e1 Macrophage Receptor with
CC      TITLE OF INVENTION:  a Colligeneous Domain and the
CC      TITLE OF INVENTION:  Polypeptide Chain Encoded by
CC      TITLE OF INVENTION:  such a Sequence
CC      NUMBER OF SEQUENCES:  2

```


CC STATE: Pennsylvania
CC COUNTRY: U.S.A.
CC ZIP: 19406-5090
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/453,117
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jervais, Herbert H.
CC REGISTRATION NUMBER: 31,171
CC REFERENCE/DOCKET NUMBER: SBC-P50338
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (610) 270-5019
CC TELEFAX: (610) 270-5090
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQENCE 451 AA; 49762 MW; 1024670 CN;
SQ

Query Match 47.7% Score 361; DB 1; Length 451;
Best Local Similarity 49.0%; Pred. No. 8.84e-27;
Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;

DB 350 VRLVGGSGPHEGRVEILHSGQWGTICDDREVRVGVCRSLGYPGQAVHKAHFGOGT 409
|||:| |||||:||||||| | : : ||||| | : |||||
QY 447 VRLMDGNNKKEGRVEVFIINGQWGTICDDGWTDKDAVAICQLGKGPARRATMAFGECK 506
|||:| |||||:||||||| | : : ||||| | : |||||
DB 410 GPIWLNEVFCFGRESSIECKIRQWGTTRACSHSDEAGVTC 449
|||:| |||||:||||||| | : : ||||| | : |||||
QY 507 GPIHDNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
|||:| |||||:||||||| | : : ||||| | : |||||

RESULT 9
ID US-08-973-145-2 STANDARD: PRT: 451 AA.
XX
AC xxxxxx
XX

Sequence 2, Application US/08973145
CC Sequence 2, Application US/08973145
CC Patent No. 5919636
CC GENERAL INFORMATION:
CC APPLICANT: Lysko, Paul G.
CC APPLICANT: Elshourbagy, Nabil A.
CC APPLICANT: Bramer, Mary E.
CC TITLE OF INVENTION: Attachment Enhanced 293 Cells
CC NUMBER OF SEQUENCES: 4
CC CURRENT APPLICATION DATA:
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham - Corporate Patents U.S.
CC STREET: Mailcode - UW2220, 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: Pennsylvania
CC COUNTRY: U.S.A.
CC ZIP: 19406-5090
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/973,145
CC FILING DATE: 26-NOV-1997
CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Elizabeth J. Hecht
CC REGISTRATION NUMBER: P-41, 824
CC REFERENCE/DOCKET NUMBER: P50338
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (610) 270-5009
CC TELEFAX: (610) 270-5090
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQENCE 451 AA; 49762 MW; 1024670 CN;
SQ

Query Match 47.7% Score 361; DB 2; Length 451;
Best Local Similarity 49.0%; Pred. No. 8.84e-27;
Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;

DB 350 VRLVGGSGPHEGRVEILHSGQWGTICDDREVRVGVCRSLGYPGQAVHKAHFGOGT 409
|||:| |||||:||||||| | : : ||||| | : |||||
QY 447 VRLMDGNNKKEGRVEVFIINGQWGTICDDGWTDKDAVAICQLGKGPARRATMAFGECK 506
|||:| |||||:||||||| | : : ||||| | : |||||
DB 410 GPIWLNEVFCFGRESSIECKIRQWGTTRACSHSDEAGVTC 449
|||:| |||||:||||||| | : : ||||| | : |||||
QY 507 GPIHDNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
|||:| |||||:||||||| | : : ||||| | : |||||

RESULT 10
ID PCT-US96-08081-2 STANDARD: PRT: 451 AA.
XX
AC xxxxxx
XX
DT
XX

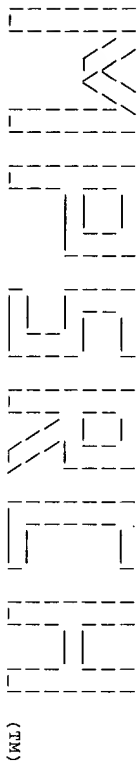
Sequence 2, Application PC/TUS9608081
CC Sequence 2, Application PC/TUS9608081
CC GENERAL INFORMATION:
CC APPLICANT: SmithKline Beecham Corporation
CC TITLE OF INVENTION: Attachment Enhanced 293 Cells
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham - Corporate Patents U.S.
CC STREET: Mailcode - UW2220, 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: Pennsylvania
CC COUNTRY: U.S.A.
CC ZIP: 19406-5090
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/08081
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jervais, Herbert H.
CC REGISTRATION NUMBER: 31,171
CC REFERENCE/DOCKET NUMBER: P50338
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (610) 270-5019
CC TELEFAX: (610) 270-5090
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQENCE 451 AA; 49762 MW; 1024670 CN;
SQ

DE Sequence 2, Application US/08794795
XX
CC Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Eishourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA: 49764 MW: 1218428 CN:
Query Match 43.1%; Score 326; DB 2; Length 495;
Best Local Similarity 48.5%; Pred. No. 3,61e-23;
Matches 49; Conservative 18; Mismatches 27; Indels 7; Gaps 5;
DB 399 VRIV-GSSNR-GRAEVYSGTWTICDDDEWQNSDAIVFCRMIGY---SKRAL-YKVGAG 452
QY 447 VRLMDGNNKKEGVFEVINGOWGTICDDGWTDPDAVVICRQLGKGPARTWAY-FGEG 505
DB 453 TGOIWLNDVQCRTSTSTMSCTKNSMGHHDCSHEDAGVGC 493
QY 506 KGPIDHNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
RESULT 14
ID US-08-794-795-6 STANDARD: PRT: 520 AA.
XX
AC xxxxxx
DE Sequence 6, Application US/08794795
XX
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:

CC APPLICANT: Eishourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA: 52658 MW: 1340662 CN:
Query Match 43.1%; Score 326; DB 2; Length 520;
Best Local Similarity 48.5%; Pred. No. 3,61e-23;
Matches 49; Conservative 18; Mismatches 27; Indels 7; Gaps 5;
DB 424 VRIV-GSSNR-GRAEVYSGTWTICDDDEWQNSDAIVFCRMIGY---SKRAL-YKVGAG 477
QY 447 VRLMDGNNKKEGVFEVINGOWGTICDDGWTDPDAVVICRQLGKGPARTWAY-FGEG 505
DB 478 TGOIWLNDVQCRTSTSTMSCTKNSMGHHDCSHEDAGVGC 518
QY 506 KGPIDHNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
RESULT 15
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.
XX
AC xxxxxx
DE Sequence 2, Application US/08200900A
XX
CC Sequence 2, Application US/08200900A
CC Patent No. 5665566
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

CC STREET: 87 Cambridgepark Drive
 CC CITY: Cambridge
 CC STATE: MA
 CC COUNTRY: USA
 CC ZIP: 02140
 CC
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/200,900A
 CC FILING DATE: 23-FEB-1994
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Meinerdt, Maureen C.
 CC REGISTRATION NUMBER: 31,544
 CC REFERENCE/DOCKET NUMBER: GI 5201-FWC
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617) 876-1170 X8574
 CC TELEFAX: (617) 876-5851
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 798 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 798 AA: 88937 MW: 3292434 CN;
 SQ
 Query Match 16.5%; Score 125; DB 1; Length 798;
 Best Local Similarity 30.6%; Pred. No. 2.50e-03;
 Matches 19; Conservative 13; Mismatches 28; Indels 2; Gaps 2;
 Db 457 VRLFNGTSSGIQVPRISIMHVACAEWNTQISDDVQCLGL-GNGNS-SVPTFTSG 514
 Qy 447 VRLMDCENKREGREYVINGQWGTICDDGWTDAVAICRQLGKGPARRIMATFEEGK 506
 Db 515 GP 516
 Qy 507 GP 508

Search completed: Mon Mar 13 10:35:28 2000
 Job time : 7 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:34:56 2000; MasPar time 6.76 Seconds
Tabular output not generated. 353.990 Million cell updates/sec

Title: >US-09-147-947-6
Description: (447-547) from US09147947A.pep (6 of 6)
Perfect Score: 757
Sequence: 1 VRLMDGENKKEGRVEVFING.....KODIGRHNCRHSEDAGVICD 101

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a:geneseqs
1:geneseqp

Statistics: Mean 28.103; Variance 102.859; scale 0.273

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Score	Query Match	Length DB	ID	Description	Pred. No.
1	757	100.0	822	1 W99087	Human serine protease	3.21e-73
2	757	100.0	875	1 W83361	Human neurotrophin	3.21e-73
3	724	95.6	761	1 W83362	Mouse neurotrophin	1.97e-69
4	724	95.6	761	1 W99088	Mouse serine protease	1.97e-69
5	391	51.7	585	1 W81363	Human cytokine regulat	1.08e-31
6	391	51.7	585	1 R36533	GP85-97 clone 18 prod.	1.08e-31
7	391	51.7	585	1 R40172	Sequence of an immunor	1.08e-31
8	391	51.7	585	1 R41359	Tumour associated 90k	1.08e-31
9	391	51.7	585	1 R40213	Sequence of a 90k tumo	1.08e-31
10	376	49.7	433	1 R05509	Scavenger receptor pro	5.02e-30
11	376	49.7	433	1 R27036	Bovine sol. scavenger	5.02e-30
12	369	48.7	489	1 W39749	Mouse macrophage recep	3.00e-29
13	369	48.7	518	1 W03561	Macrophage receptor w1	3.00e-29
14	366	48.3	451	1 W19708	Macrophage scavenger r	6.45e-29
15	366	48.3	451	1 W64591	Human SRCR protein.	6.45e-29
16	361	47.7	451	1 R27035	Human sol. scavenger r	2.31e-28
17	361	47.7	451	1 W08077	Type I macrophage scav	2.31e-28
18	361	47.7	451	1 R40802	Human scavenger recept	2.98e-28
19	360	46.9	451	1 R21512	Human scavenger recept	1.07e-27
20	355	46.2	666	1 W64590	Human SRCR protein fra	3.80e-27
21	350	45.7	1290	1 W07609	Rat von Ebner's gland	1.05e-26
22	346	45.7	574	1 R40784	Sequence of cyclopallil	1.05e-26
23	346	45.7	574	1 W81362	Mouse cytokine regulat	1.05e-26

ID	Score	Query Match	Length DB	ID	Description	Pred. No.
24	332	43.9	347	1 Y13369	Amino acid sequence of	3.67e-25
25	332	43.9	347	1 W68200	Human scavenger recept	3.67e-25
26	332	43.9	347	1 W64537	Human liver cell clone	3.67e-25
27	326	43.1	495	1 W39747	Human macrophage recep	1.68e-24
28	326	43.1	520	1 W39748	Human macrophage recep	1.68e-24
29	326	43.1	520	1 Y04374	Human macrophage recep	1.68e-24
30	326	43.1	520	1 Y04374	Human macrophage recep	1.68e-24
31	312	16.5	798	1 R57283	Bovine enterokinase.	2.25e-03
32	312	16.5	798	1 R57283	Bovine enterokinase.	2.25e-03
33	85	11.2	755	1 W81365	Human CD5 for use in T	1.14e+00
34	82	10.8	356	1 W46917	Human prothrombin conve	1.06e+01
35	80	10.6	380	1 W54350	Amino acid sequence of	1.92e+01
36	80	10.6	380	1 W54350	Amino acid sequence of	1.92e+01
37	79	10.4	1167	1 R54073	CAMP-dependent protein	2.84e+01
38	79	10.4	1167	1 R54073	CAMP-dependent protein	2.84e+01
39	79	10.4	1167	1 W87632	Cyert4.	2.84e+01
40	77	10.2	1167	1 W17700	Cyert4.	2.84e+01
41	77	10.2	1162	1 W06829	Bacillus thuringiensis	3.45e+01
42	76	10.0	198	1 R59841	Major neutralizing ant	5.07e+01
43	76	10.0	198	1 R59841	Major neutralizing ant	5.07e+01
44	76	10.0	311	1 R92115	Sequence of a region o	6.14e+01
45	76	10.0	311	1 R59843	ApoE4L protease.	6.14e+01

ALIGNMENTS

ID	Score	Query Match	Length DB	ID	Description	Pred. No.
1	757	100.0	822	1 W99087	Human serine protease	3.21e-73
2	757	100.0	875	1 W83361	Human neurotrophin	3.21e-73
3	724	95.6	761	1 W83362	Mouse neurotrophin	1.97e-69
4	724	95.6	761	1 W99088	Mouse serine protease	1.97e-69
5	391	51.7	585	1 W81363	Human cytokine regulat	1.08e-31
6	391	51.7	585	1 R36533	GP85-97 clone 18 prod.	1.08e-31
7	391	51.7	585	1 R40172	Sequence of an immunor	1.08e-31
8	391	51.7	585	1 R41359	Tumour associated 90k	1.08e-31
9	391	51.7	585	1 R40213	Sequence of a 90k tumo	1.08e-31
10	376	49.7	433	1 R05509	Scavenger receptor pro	5.02e-30
11	376	49.7	433	1 R27036	Bovine sol. scavenger	5.02e-30
12	369	48.7	489	1 W39749	Mouse macrophage recep	3.00e-29
13	369	48.7	518	1 W03561	Macrophage receptor w1	3.00e-29
14	366	48.3	451	1 W19708	Macrophage scavenger r	6.45e-29
15	366	48.3	451	1 W64591	Human SRCR protein.	6.45e-29
16	361	47.7	451	1 R27035	Human sol. scavenger r	2.31e-28
17	361	47.7	451	1 W08077	Type I macrophage scav	2.31e-28
18	361	47.7	451	1 R40802	Human scavenger recept	2.98e-28
19	360	46.9	451	1 R21512	Human scavenger recept	1.07e-27
20	355	46.2	666	1 W64590	Human SRCR protein fra	3.80e-27
21	350	45.7	1290	1 W07609	Rat von Ebner's gland	1.05e-26
22	346	45.7	574	1 R40784	Sequence of cyclopallil	1.05e-26
23	346	45.7	574	1 W81362	Mouse cytokine regulat	1.05e-26

At the end of the available but from

KW emptysema; bronchitis.
 OS Homo sapiens.
 PN WO9849332-A1.
 PD 05-NOV-1998.
 PR 24-APR-1998: IB0625.
 PF 26-APR-1997: CH-000966.
 PI (SOND/) SONDEREGGER P.
 DR Sonderegger P;
 WI: 99-009438/01.
 N-PSDB: V72589.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 20-24; 50pp: English.
 CC The present sequence represents human neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases; e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emptysema or
 CC bronchitis).
 SQ Sequence 875 AA;

Query Match 100.0%; Score 757; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 3,21e-73;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 500 VRLMDGNKKEGVEVFVINGQWGTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 559
 QY 447 VRLMDGNKKEGVEVFVINGQWGTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 506
 DB 560 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYICD 600
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYICD 547

RESULT 3
 ID W83362 standard; Protein: 761 AA.
 AC W83362.
 DT 17-FEB-1999 (first entry)

DE Mouse neurotysin.
 OS Mouse; neurotysin; tumour inhibition; neurological disease;
 OS neurodegeneration; gene therapy; drug development; stroke; brain injury;
 OS neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 OS epilepsy; hypoxia; ischaemia; nerve transection; neovascularisation;
 OS emptysema; bronchitis.
 OS Mus musculus.
 PN WO9849332-A1.
 PD 05-NOV-1998.
 PR 24-APR-1998: IB0625.
 PF 26-APR-1997: CH-000966.
 PI (SOND/) SONDEREGGER P.
 DR Sonderegger P;
 WI: 99-009438/01.
 N-PSDB: V72590.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 29-32; 50pp: English.
 CC The present sequence represents mouse neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases; e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emptysema or
 CC bronchitis).
 SQ Sequence 761 AA;

Query Match 95.6%; Score 724; DB 1; Length 761;
 Best Local Similarity 92.1%; Pred. No. 1.97e-69;
 Matches 93; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 386 IRLVDGNKKEGVEVFVINGQWGTICDDGWTDKHAAVICRQLGYGPARARTMAFGECK 445
 QY 447 VRLMDGNKKEGVEVFVINGQWGTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 506
 DB 446 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYICD 486
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYICD 547

RESULT 4
 ID W99088 standard; Protein: 761 AA.
 AC W99088.
 DT 13-MAY-1999 (first entry)

DE Mouse serine protease BSSP-3.
 OS Serine protease; BSSP-3; brain tissue.
 OS Mus sp.
 PN WO9905290-A1.
 PD 04-FEB-1999.
 PR 24-JUL-1998: J03324.
 PF 24-JUL-1997: JP-213969.
 PA (SUNR) SUNTOY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR WPI: 99-142942/12.
 N-PSDB: X19027.
 PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1: Page 51-54; 69pp: Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transformants may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 761 AA;

Query Match 95.6%; Score 724; DB 1; Length 761;
 Best Local Similarity 92.1%; Pred. No. 1.97e-69;
 Matches 93; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 386 IRLVDGNKKEGVEVFVINGQWGTICDDGWTDKHAAVICRQLGYGPARARTMAFGECK 445
 QY 447 VRLMDGNKKEGVEVFVINGQWGTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 506
 DB 446 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYICD 486
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYICD 547

RESULT 5
 ID W81363 standard; Protein: 585 AA.

AC W81363;
 DT 12-APR-1999 (first entry)
 DE Human cytokine regulatory factor CYTRF.
 CC Cytokine regulatory factor; cyclophilin C-associated protein;
 KW CYTRF; CycAP; antiinflammatory; inflammation; autoimmune disease;
 KW multiple sclerosis; rheumatoid arthritis; ankylosing spondylitis;
 KW meningitis; herpes encephalitis; viral meningoencephalitis;

KW viral hepatitis; graft rejection; graft versus host disease;
 OS allergy; asthma; diagnosis; therapy; human.
 FH Homo sapiens.
 FT Misc_difference 257 /note= "encoded by CGT"
 FN W09856819-A1.
 PD 17-DEC-1998.
 PD 12-JUN-1998; U12345.
 PR 13-JUN-1997; US-049578.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Conboy IB Jones PP.
 DR WPI: 99-080890/07.
 DR N-PSDB: V68491.
 CC Reduction of pro-inflammatory cytokines in a T cell population -
 using cytokine regulatory factor (CYRF), useful in the diagnosis
 and treatment of pro-inflammatory conditions
 Disclosure: Page 47-51; 77pp; English.
 CC This is the amino acid sequence of human cytokine regulatory factor
 (CYRF), also known as cyclophilin C-associated protein (CyCAP).
 CC Methods are provided for the modulation of cytokine production by T
 CC helper cells using CYRF (especially human CYRF). A pharmaceutical
 CC composition comprising CYRF as an active agent is administered in
 CC vitro or in vivo, and can act on mature, committed Th type T cells
 CC to decrease the production of pro-inflammatory cytokines, or to skew
 CC the commitment of precursor T helper cells to Th1 or a Th1 cytokine
 CC profile. The administration of CYRF is useful in the diagnosis and
 CC treatment of autoimmune diseases (e.g. multiple sclerosis,
 CC rheumatoid arthritis, and degenerative joint diseases including
 CC ankylosing spondylitis and reactive arthritis), and inflammation
 CC caused by bacterial, viral or parasitic infection (e.g. meningitis,
 CC herpes encephalitis, viral meningoencephalitis and viral hepatitis),
 CC including response to vaccination (e.g. vaccines containing rabies,
 CC varicella zoster, and measles) and skin sensitivity (e.g. graft
 CC rejection and graft vs host disease). CYRF signaling and related
 CC pathways are also useful for modelling and screening new
 CC pharmacological agents. The administration of agents that block
 CC CYRF action (e.g. antibodies) are useful in reducing unwanted
 CC allergic responses, especially in asthma.
 SO Sequence 585 AA;

Query Match 51.7%; Score 391; DB 1; Length 585;
 Best Local Similarity 52.0%; Pred. No. 1.08e-31;
 Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

24 MRLADGATNGRVEIFRGQMGTCVCDNLMDLTDASVCRALGFENATQALGRAFGGGS 83
 447 VRLMDGKRRKGRVEIFNGQMGTCVCDGWDKDAVAICRDLGKGRPARFTMAYFGGSK 506
 DB 84 GPIIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVVC 123
 507 GPIHYDNVCKTGNERSLADCKIKODIGRNRCHSESDAGVVC 546

RESULT 6
 ID R36533 standard; Protein; 585 AA.
 AC R36533:
 DE 19-AUG-1993 (first entry)
 DF gp85-97 clone 18 prod.
 KW Glycoprotein; lectin; Mac-2; PHA; lymphocytes; leucoagglutinating;
 KW phytohemagglutinin; cancer; amplification; PCR.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 1..18
 FT /note= "leader sequence"
 FT protein 19..585
 FT /note= "mature gp85-97"
 FN W09308215-A.
 FN 29-APR-1993.
 PD 15-OCT-1992; U08878.
 PR 16-OCT-1991; US-777121.
 PR 15-OCT-1992; US-961404.
 PA (CERTU) CERTUS ONCOLOGY CORP.

PI Casipit CL, Halenbeck R, Kolts KE, Taylor EM, Wang AM;
 DR WPI: 93-152424/18.
 PT New glyco:protein complex binding to human lectin Mac-2 - also
 PT interferes with PHA activation of lymphocytes for treating and
 PT preventing cancer, infectious diseases, etc.
 PS Claim 1; Page 41; 57pp; English.
 CC SK-BR-3 gp97 recovered in partially proteolysed form was denatured
 CC and reduced and the 97 and 70 kD mols. were purified using size
 CC exclusion HPLC in 0.1 percent SDS. The 97 and 70 kD mols. were
 CC digested with Lys-C protease and the resulting peptides purified and
 CC sequenced. The N-terminal sequence of gp97 was used to design
 CC degenerate primers for use in a PCR reaction on SK-BR-3 mRNA. PCR
 CC reactions using primer 2 in combination with primers 1 and 3 gave prods.
 CC of 97 and 121 bp. Further PCR using primers 2, 3 and 4 yielded DNA
 CC sequences of 740 and 765 respectively. A DNA sequence was obt'd. from
 CC the PCR prods. Two additional oligonucleotide sequences were
 CC synthesized based on this sequence and used to probe a THP-1 CDNA
 CC library to obtain the full-length CDNA sequence that encodes SK-BR-3
 CC gp97. Both strands of clone 218 were sequenced. It encodes a novel
 CC glycoprotein binds the Mac-2 lectin and interferes with PHA activation
 CC of lymphocytes. It can be used for treating or preventing diseases that
 CC result from binding of a disease-causing agent to the cell surface of a
 CC target cell. The gp. can be used in treatment of cancers, partic.
 CC breast cancer. See also R36532.
 SO Sequence 585 AA;

Query Match 51.7%; Score 391; DB 1; Length 585;
 Best Local Similarity 52.0%; Pred. No. 1.08e-31;
 Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRGQMGTCVCDNLMDLTDASVCRALGFENATQALGRAFGGGS 83
 447 VRLMDGKRRKGRVEIFNGQMGTCVCDGWDKDAVAICRDLGKGRPARFTMAYFGGSK 506
 DB 84 GPIIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVVC 123
 507 GPIHYDNVCKTGNERSLADCKIKODIGRNRCHSESDAGVVC 546

RESULT 7
 ID R40172 standard; Protein; 585 AA.
 AC R40172:
 DE 08-FEB-1994 (first entry)
 DE Sequence of an immunoregulin (IR)-95 polypeptide.
 KW Tumour associated antigen; breast cancer; cell line CG-5;
 KW ovarian cancer; immunoregulin-95; IR-95.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 1..17
 FT /label= signal
 FT region 24..125
 FT /label= SRCR homology region
 FT modified_site 71
 FT /label= glycosylation site
 FT /note= "see also AAs 125,192,362,398,551,580"
 FN W09316180-A.
 FN 19-AUG-1993.
 PD 17-FEB-1993; E00382.
 PR 17-FEB-1992; IT-R0100.
 PA (PLAC) MAX PLACK GES FORDERUNG WISSENSCHAFTEN.
 PA (UYCH-) UNIV CHEITI ANNUNIO G D.
 PI Azam M, Iacobelli S, Natoli C, Sures I, Ullrich A;
 DR WPI: 93-372884/34.
 DR N-PSDB: Q46888.
 PT Recombinant DNA - encoding tumour associated antigen.
 PT Immuno-regulin-95
 PS Disclosure: Fig 1; 69pp; English.
 CC Immunoregulin-95 is a 90k tumour-associated antigen purified from
 CC the culture fluid of the human breast cancer cell line, CG-5, the
 CC serum of a breast cancer patient, or the ascitic fluid from an
 CC ovarian cancer patient. The native antigen, which has a mol. wt. of
 CC 95 kD, is present as a high mol. wt. complex. Homology in the region
 CC of AAs 35-80 of the 90k antigen is found with type I macrophage

CC scavenger receptor; sea urchin speract receptor; and human
CC lymphocyte glycoprotein TI/Len-1.
SQ Sequence 585 AA;

Query Match 51.7%; Score 391; DB 1; Length 585;
Best Local Similarity 52.0%; Pred. No. 1.08e-31;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRYGOMGTVCNLMDLTDASVYCRALGFENATQALGRAAFGQS 83
Y 447 VRLMDGKNGKRGREVEIFNGOMGTICDDGTDKAAVTCRQLGYGPARARTMAFGECK 506
DB 84 GPIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVC 123
Y 507 GPIHDVNYKCTGNERSLADCKIKODIGRHNCRHSEDAVC 546

RESULT 8
R41359 standard; Protein; 585 AA.

R41359; 03-MAR-1994 (first entry)
Tumour associated 90K antigen.
Antigen: cancer; inflammation; autoimmune disease; viral infection.
Homo sapiens.

FT Key peptide Location/Qualifiers
FT 1..18 /label= Signal peptide.
FT 24..125 /note= "Region homologous with sea urchin speract receptor"

FT modified_site 69
FT /note= "Potential asparagine linked N-glycosylation site"
FT modified_site 125
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 192
FT /note= "Potential asparagine linked N-glycosylation site"
FT modified_site 362
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 398
FT /note= "Potential asparagine linked N-glycosylation site"
FT modified_site 551
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 580
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 585
FT /note= "Potential asparagine linked N-glycosylation site"

MO9317119-A.
02-SEP-1993.
15-FEB-1993; E00379.
PR 17-FEB-1992; IT-RM0099.
PA (UYCH-) UNIV CHIETI ANNUNZIO G D.
PA (UYNY) UNIV NEW YORK STATE.
PI Iacobelli S, Natoli C, Schlessinger J;
DR WPI: 93-288423/36.

DR N-PSDB; Q46161.
PT 90K tumour-associated antigen - purified from serum of breast cancer patient or ascitic fluid from ovarian cancer patient
PS Claim 1; Figure 1; 73pp; English.
CC The purified 90K antigen or its antigenic determinant containing fragment can be assayed to diagnose a disorder, such as cancer (especially breast or ovarian cancer) or a viral infection, inflammation, autoimmune disease and/or arthritis in a patient.
CC It may also be used in the preparation of an agent for the treatment of these disorders.
CC Sequence 585 AA;

Query Match 51.7%; Score 391; DB 1; Length 585;
Best Local Similarity 52.0%; Pred. No. 1.08e-31;

Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRYGOMGTVCNLMDLTDASVYCRALGFENATQALGRAAFGQS 83
Y 447 VRLMDGKNGKRGREVEIFNGOMGTICDDGTDKAAVTCRQLGYGPARARTMAFGECK 506
DB 84 GPIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVC 123
Y 507 GPIHDVNYKCTGNERSLADCKIKODIGRHNCRHSEDAVC 546

RESULT 9
R40213 standard; Protein; 585 AA.

R40213; 03-FEB-1994 (first entry)
DE Sequence of a 90K tumour-associated antigen, IR-95, capable of binding to SP-2 (Accession No. I-1083).
DE Tumour-associated antigen; breast cancer cell line CG-5;
KM 90 K antigen.
OS Homo sapiens.

FT Key peptide Location/Qualifiers
FT 1..18 /label= signal
FT 24..125 /label= SCRC homology region

FT binding_site 69
FT /label= potential glycosylation site
FT /note= "see also AAs 125,192,363,398,551,580"

MO9316181-A.
19-AUG-1993.
PR 17-FEB-1993; E00385.
PA (UYCH-) UNIV CHIETI ANNUNZIO G D.
PA (UYNY) UNIV NEW YORK STATE.
PI Iacobelli S, Natoli C, Schlessinger J;
DR WPI: 93-272885/34.

DR N-PSDB; Q48170.
PT New 90K tumour-associated antigen, IR-95 - is for use in diagnosis and therapy of cancer, HIV and auto-immune diseases
PS Claim 2; pages 45-48; 68pp; English.

CC 90K antigen has an apparent molecular weight of approx. 95 kd. It is a tumour associated antigen and is elevated in the serum of patients with cancer and also in patients with HIV. It reacts with Mab SP-2 which was prod. by immunising mice with proteins that had been released into tissue culture fluid by human MCF-7 breast cancer cells. Mab SP-2 cell line is deposited at the Institut Pasteur, Paris. Accession number I-1083. 90K is also present in normal subjects. It is purified from the culture fluid of the human breast cancer cell line; CG-5. N-terminal sequencing of the 90K antigen was used to design a 'guesser' nucleotide sequence (Q48171) as a probe to screen a lambda-gt10 library pred. from MCF7 polyA+ RNA.
CC The complete nucleotide sequence of isolated clones is given in Q48170.
CC Sequence 585 AA;

Query Match 51.7%; Score 391; DB 1; Length 585;
Best Local Similarity 52.0%; Pred. No. 1.08e-31;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRYGOMGTVCNLMDLTDASVYCRALGFENATQALGRAAFGQS 83
Y 447 VRLMDGKNGKRGREVEIFNGOMGTICDDGTDKAAVTCRQLGYGPARARTMAFGECK 506
DB 84 GPIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVC 123
Y 507 GPIHDVNYKCTGNERSLADCKIKODIGRHNCRHSEDAVC 546

RESULT 10
R05509 standard; Protein; 453 AA.

R05509; 23-OCT-1990 (first entry)
DE Scavenger receptor protein with affinity for acylated low density

507 GPIHVDNVKCTGNERSLADCIKQDIGRHNCRHSEDAGVIC 546

RESULT	13	
ID	W03561	standard; Protein; 518 AA.
AC	W03561:	
DT	22-JUN-1997	(first entry)
DE	Macrophage receptor with collagenous domain (MARCO).	
KW	Macrophage receptor-collagenous; MARCO; immune defense;	
OS	Mus sp.	
EH	Key	
FT	domain	Location/Qualifiers
FT		1..49
FT	/label= Domain-I	
FT	/note= "hydrophilic domain not containing a signal peptide"	
FT	domain	50..74
FT	/label= Domain-II	
FT	/note= "hydrophobic domain"	
FT	domain	75..149
FT	/label= Domain-III	
FT	/note= "hydrophilic domain"	
FT	modified_site	87..89
FT		/label= Glycosylation
FT	modified_site	/note= "putative N-glycosylation site"
FT		138..140
FT	/label= Glycosylation	
FT	/note= "putative glycosylation site"	
FT	domain	150..419
FT	/label= Domain-IV	
FT	/note= "collagenous structure characterised by 89 Gly-X-X repeats with one interruption (Ala-Glu-Lys at 174-176)"	
FT	domain	420..518
FT		/label= Domain-V
FT		/note= "globular C-terminal domain"
PN	W0626219-AI.	
PD	29-AUG-1996.	
PE	19-FEB-1996; FI0091.	
PF	21-FEB-1995; US-392367.	
PA	(ELDM/) ELDM A O.	
PA	(KANG/) KANGAS M.	
PA	(TRYG/) TRYGVASON K.	
PI	Elomaa O, Kangas M, Trygvason K;	
DR	WPI: 96-402322/40.	
DR	N-PSDB: T39340.	
PT	New macrophage receptor, MARCO, that binds bacteria but not yeast	
PT	and related DNA, anti-sense probes and antibodies, involved in	
PT	immune defence and phagocytosis.	
PT	Claim 10; Fig 2A-B; 62pp; English.	
CC	A novel macrophage receptor with a collagenous domain, named MARCO	
CC	(W03561), is a membrane-bound trimeric protein which is expressed	
CC	strongly after birth in a subset of macrophages in mouse spleen and	
CC	lymph nodes. It binds to bacteria but not to yeast, and probably	
CC	plays a role in immune defence and/or phagocytosis. The amino acid	
CC	sequence of MARCO was deduced from DNA (see also T39340) obtd. from	
CC	a mouse macrophage cDNA library. Recombinant MARCO can be produced	
CC	in transformed host cells. It is used to raise antibodies useful	
CC	in detecting MARCO or interfering with its function.	
CC	Sequence 518 AA;	

	Query Match	48.7%	Score 369;	DB 1;	Length 518;
	Best Local Similarity	54.0%;	Pred. No. 3,00e-29;		
	Matches	54;	Conservative	17;	Mismatches 24; Indels 5; Gaps 2
D b	423	VRIWGSTR--GRAEVYNNEMGTCDDDDMNNAATVCRMLGY---SRGRALSSYGGS	477		
		: : : : : : : : :			
O y	447	VRLDMEKKKGGRAVEFTNGQWGTICDDGWTDKDAAYCROLGYGPARTMAFYEGCK	506		
		: : : : : : : : :			
D b	478	GNIWLDNANCGTESSLWDCCSKNSWGNCHNEADAGEC	517		
		: : : :			
O y	507	GPIHVDNVKCTGNERSLADCIKODIGRINCHHSBEDAGYIC	546		

RESULT 14
ID W19708 standard; Protein: 451 AA.
AC W19708:
DT 19-AUG-1997 (first entry)
DE Macrophage scavenger receptor protein.
KW Macrophage scavenger receptor protein; MSRP; human; lipoteichoic acid;
KW LTA; Gram-positive bacteria; cell wall; atherosclerosis; host defence;
KW septicæmia; inhibitor; complement activation; cytokine release; therapy.
KW nitric oxide production; bacterial infection; septic shock.
OS Homo sapiens.
PN US5624904-A.
PD 29-APR-1997.
PE 17-NOV-1993.
PR 17-NOV-1993; US-154365.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Joiner KA, Krieger M;
DR MPI: 97-258236/23.
DR N-PSDB: T68796.
PT Treatment of septicæmia caused by Gram-positive bacteria - by
PT administration of macrophage scavenger receptor protein
PS Disclosure: Column 23-26: 16pp: English.
CC This sequence represents the human macrophage scavenger receptor protein
CC (MSRP). MSRP binds specifically to the lipoteichoic acid (LTA) residues
CC on the Gram-positive bacterial cell wall. MSRP exhibit unusually broad
CC binding specificity for polyanionic ligands, and have been implicated in
CC atherosclerosis and a variety of host defence functions. This protein can
CC be used in the method of the invention. The method of the invention is
CC for the treatment of septicæmia caused by Gram-positive bacteria. The
CC method comprises administration of a MSRP in an amount sufficient to
CC inhibit complement activation, cytokine release or nitric oxide
CC production induced by LTA released by the Gram-positive bacteria. In
CC addition to treating Gram-positive septic shock, the MSRP can be used to
CC screen for other compounds for treating Gram-positive septic shock. MSRP
CC can also be used to purify, label or detect LTA or LTA-containing cells.
CC MSRP, active MSRP fragments, anti-MSRP antibodies or other compounds that
CC inhibit binding of Gram-positive bacteria to MSRP can be used to treat
CC pathologies such as septicæmia, Gram-positive bacterial infection,
CC Gram-positive septicæmia or Gram-positive or Gram-negative septic shock.
SQ Sequence 451 AA;

Query Match:	48.3%	Score	366	DB	1	Length	451
Best Local Similarity:	49.0%	Pred.	No. 6.45e-29;				
Matches	49;	Conservative	17;	Mismatches	34;	Indels	0; Gaps 0;
Db	350	VRLVGGSGPHEGRVEITLHSGWGTICDDNNEPRVGVCVCSLTGPGVQAVHKAAHFGGCT	409				
		: : : : : :					
OY	447	VRLMDGENKKEGRVEVFINGMGITCIDDGWTDKAALVTCQLQSLKGPARARMTAFYGESEK	506				
Db	410	GPIMLNEVFCFGRSIEIECKIRQMGTRACSHSEDACVTC	449				
OY	507	GPIHDVNKCTGNERSLADCIODIGRNCRHSDSDACVTC	546				

RESULT 15
ID W64591 standard; Protein; 1785 AA.
AC W64591;
DT 23-Oct-1998 (first entry)
DE Human SRCR protein.
KW Sarcovagel receptor, cysteine rich domain, SRCR, diagnosis; treatment;
KW nervous system, medullo-blastoma; glioma; breast; detection;
KW autoantibody; ss.
OS Homo sapiens.
PN W09830687-A2.
PD 16-JUL-1998.
PE 09-JAN-1998; D00096.
PR 18-JUL-1997; DE-030997.
PR 09-JAN-1997; DE-000519.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Mollenhauer J, Poustka A;
DR WPI; 98-399136/34.
DR N-PSDB; V49652.

THIS PAGE BLANK (USPTO)


```

OY 507 GPIHYDNVAKCTGNERSLADCIKODIGRHNCRHSEDAGVICD 547

RESULT 2
ENTRY 114893 #type complete
TITLE scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name purple urchin
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS
REFERENCE T14893
#authors Pancer, Z.; Rast, J.P.; Davidson, E.H.
#journal Immunogenetics (1999) 49:773-786
#title Origins of immunity: transcription factors and homologs of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes.
#accession T14893
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-2153 #label PAN
##cross-references EMBL:AF064259; NID:g4165052; PID:g4165053; PIDN:AD08654.1

GENETICS
SUMMARY SRCR12
#gene #length 2153 #molecular_weight 226556 #checksum 3510
#gene #length 2153 #score 466; DB 2; length 2153;
#best_local_similarity 60.0%; Pred. No. 4,71e-66;
#matches 60; Conservative 13; Mismatches 27; Indels 0; Gaps 0;

Db 1837 VALFGNNHNEGVYEFPLAGTWTGTCDDGDIDIASVYCHLQGSARATNAYFGGGE 1896
OY 447 VRLMDGKREKGEVVFINGQWGTICDDGTDKRAAAYICROLGYKGPARTMAYFGEK 506
Db 1897 GPIYMDGVSCDGEDPLSMCHHAGIFGHNCGHODAVWC 1936
OY 507 GPIHYDNVAKCTGNERSLADCIKODIGRHNCRHSEDAGVIC 546

RESULT 3
ENTRY A47161 #type complete
TITLE Mac-2-binding glycoprotein precursor - human
ALTERNATE_NAMES 90K tumor-associated protein precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 05-May-1995 #sequence_revision 05-May-1995 #text_change 16-Jul-1999

ACCESSIONS
REFERENCE A47161
#authors Kochs, K.; Taylor, E.; Halenbeck, R.; Casipit, C.; Wang, A.
#journal J. Biol. Chem. (1993) 268:14245-14249
#title Cloning and characterization of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain.
#cross-references MUID:93300818
#accession A47161
##status preliminary
##molecule_type mRNA
##residues 1-585 #label KOT
##cross-references GB:L13210; NID:g307152; PID:g307153
REFERENCE A53687
#authors Ullrich, A.; Sures, I.; D'Egidio, M.; Jallat, B.; Powell, T.J.; Herbst, C.; Dreps, A.; Adam, M.; Rubinstein, M.; Nattoli, C.; Shawver, L.K.; Schlessinger, J.; Iacobelli, S.
#journal J. Biol. Chem. (1994) 269:18401-18407
#title The secreted tumor-associated antigen 90K is a potent immune stimulator.
#cross-references MUID:94308070
#accession A53687
##status preliminary
##molecule_type mRNA
#residues 1-585 #label ULL

```

```

REFERENCE ##cross-references GB:X79089; NID:g483473; PIDN:CAAS6599.1; PID:g483474
S32384
#authors Iacobelli, S.; Buccì, I.; d'Egidio, M.; Giuliani, C.; Natoli,
C.; Tinari, N.; Rubinstein, M.; Schlessinger, J.
#journal FEBS Lett. (1993) 319:59-65
#title Purification and characterization of a 90 kDa protein
released from human tumors and tumor cell lines.
#cross-references MIMD:9320227
#accession S32384
##molecule_type Protein
#residues 19-24,'C',26-40 #label IAC
REFERENCE PC2211
#authors Inohara, H.; Raz, A.
#journal Biochem Biophys Res Commun. (1994) 201:1366-1375
#title Identification of human melanoma cellular and secreted
ligands for galactin-3.
#cross-references MIMD:94296411
#accession PC2211
##molecule_type Protein
#residues 19-38 #label INO
#note the reported molecular weight of this protein was 98K
#accession PC2212
##molecule_type Protein
#residues 19-26 #label IN2
#note the reported molecular weight of this protein was 70K
REFERENCE A41005
#authors Rosenberg, I.; Cherayil, B.J.; Isselbacher, K.J.; Pillai, S.
#journal J. Biol. Chem. (1991) 266:18731-18736
#title Mac-2-binding glycoproteins. Putative ligands for a cytosolic
beta-galactoside lectin.
#cross-references MIMD:92011634
#accession A41005
##molecule_type Protein
#residues 19-22,'G',24-27,'XX',30 #label ROS
#note the reported molecular weight of this protein,
designated Mac-2-binding glycoprotein 2 (M2BP-2), was
70k; the amino-terminal sequence of a related form of
98k, designated M2BP-1, could not be determined
extracellular protein; glycoprotein
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS extracellular protein; glycoprotein
FEATURES
1-18
21-124
SUMMARY #domain signal sequence #status predicted #label SIG\
#domain scavenger receptor cysteine-rich domain homology
#label SNC
#length 585 #molecular-weight 65330 #checksum 2263
Query Match 51.7%; Score 391; DB 2; Length 585;
Best Local Similarity 5.0%; Pred. No. 7,87e+68;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;
Db 24 MLADGGATNGRVEYIFRGQWGTCDNLMDLTDSAVCRALGFENATQALGRAFGGS 83
::: ||| ::|||::| |::|::| |::|::| |::|::| |::|::| |::|::|
Oy 447 VRLMDGEMKREGREYEVINGOWCTICDDGWTKDAVAIVCRQLGYGPAPARMTAYFEGSK 506
||||::|::| |::|::| |::|::| |::|::| |::|::| |::|::| |::|::|
Db 84 GPMLDEVCTGTPLASLADCKSLGTLKSCRCRHEDAGVC 123
||||::|::| |::|::| |::|::| |::|::| |::|::| |::|::| |::|::|
Oy 507 GPIHDNVKCTGNERSLADCIKODIGRNHCNSHSEDAVIC 546
||||::|::| |::|::| |::|::| |::|::| |::|::| |::|::| |::|::|
RESULT 4
ENTRY 508276 #type complete
TITLE LDL receptor I, macrophage - bovine
ALTERNATE_NAMES macrophage scavenger receptor type I
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
20-Sep-1999
ACCESSIONS 508276
REFERENCE S08276
#authors Kodama, T.; Freeman, M.; Rohrer, L.; Zabrecky, J.;
Matsudaira, P.; Krieger, M.
Nature (1990) 343:531-535
Title Type I macrophage scavenger receptor contains alpha-helical
and collagen-like coiled coils.

```



```
#residues      1-1149 ##label RES  
#cross-references EMBL:D22971; NID:g312147; PID:g312148  
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology  
FEATURE  
43-147  
  
    151-254          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC1\  
    258-361          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC2\  
    365-468          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC3\  
    470-573          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC4\  
    608-711          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC5\  
    744-847          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC6\  
    849-953          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC7\  
    954-1057        #domain scavenger receptor cysteine-rich domain homology  
                #label SRC8\  
            FARY     #length 1149 #molecular-weight 124328 #checksum 487  
  
Query Match           50.2% Score 380; DB 2; Length 1149;  
Best Local Similarity 51.0%; Pred. No.3,4de+65;  
Matches   51; Conservative 19; Mismatches 30; Indels 0; Gaps 0;
```

Dd

```
747 LRLVNGGRGRCAGVEYIEHSGSMTICDDSDLDPAHYVCROLGCGEAINATGSAPFEGT 806  
:::| : |:::: | ::|| :: || ::|| ::|| ::|| ::|| ::|| ::||  
Qy 447 VLMDENKKEGVVFINGQMGTICDGDGTADDAVICOLQKGFRARNTMAFYDEGK 506  
:::| : |:::: | ::|| :: || ::|| ::|| ::|| ::|| ::|| ::||  
Db 807 GPIWDEMCKNGEKSRIRIWCCHSHGWGCQNCRHKEDAVIC 846  
||||:||||:| : | : |:||||| |||||  
Qy 507 GPITVDNVAKCTGNERSLCDICIKODIGRHCNRSHEADGAVIC 546  
||||:||||:| : | : |:||||| |||||
```

RESULT 8

```
ENTRY              I38004             #type complete  
TITLE              M130 antigen (cytosolic variant 1) - human  
ORGANISM            Homo sapiens #common_name man  
DATE                17-May-1996 #sequence_revision 17-May-1996 #text_change  
                   07-Feb-1997  
ACCESSIONS         I38004; S36078  
REFERENCE  
AUTHORS            Law,S.K.; Micklem,K.J.; Shaw,J.M.; Zhang,X.P.; Dong,Y.;  
                  Willis,A.C.; Mason,D.Y.  
JOURNAL            Eur.J.Immunol.(1993) 23:2320-2325  
#title            A new macrophage differentiation antigen which is a member of  
                 The scavenger receptor superfamily.
```

Cross-references MOID:93380506

```
I38004  
status             preliminary; translated from GB/EMBL/DDBJ  
classification     ##molecule_type mRNA  
KEYWORDS            ##residues 1-1151 ##label RES  
                    #cross-references EMBL:D22969; NID:g312143; PID:g312144  
                    #superfamily scavenger receptor cysteine-rich domain homology  
CYTO SOL  
43-147             #domain scavenger receptor cysteine-rich domain homology  
                #label SRC1\  
    151-254          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC2\  
    258-361          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC3\  
    365-468          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC4\  
    470-573          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC5\  
    575-678          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC6\  
    711-814          #domain scavenger receptor cysteine-rich domain homology  
                #label SRC7\
```

Query Match	50.2%	Score 380;	DB 2;	Length 1151;
Best Local Similarity 51.0%;	Pred. NO. 3,48e-65;			
Matches 51;	Conservative 19;	Mismatches 30;	Indels 0;	Gaps 0;
816-920	#domain scavenger receptor cysteine-rich domain	homology		
921-1024	#label SRC8\	#domain scavenger receptor cysteine-rich domain	homology	
SUMMARY	#label SRC9	#domain scavenger receptor cysteine-rich domain	homology	
	#length 1151	#molecular-weight 124820	#checksum 1481	
Db	714	LRVNGGRCAGRYEIIHESGWTICDDSDWLDSDAHVVCRLGCGEAINATGSAHFEGGT	773	
QY	447	VRLDGKKRGREYFINGOWGICDDGTBDKDAVIRQLGTYGPARFARTMAVFGSGK	506	
Db	774	GPWLDEKCKNGKESRIMWCHSHGQOQCRKEDAGVIC	813	
QY	507	GPVHDVVKCTGNERSLADCIKODIGRHNCRHSDDAGVIC	546	
RESULT	9			
ENTRY	I38005	#type complete		
TITLE	M130 antigen (cytosolic variant 2) - human			
ORGANISM	#formal_name Homo sapiens #common_name man			
DATE	17-May-1996 #sequence_revision 17-May-1996 #text_change 07-Feb-1997			
ACCESSIONS	I38005; S56079			
REFERENCE	I38003			
#authors	Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.			
#journal	Eur. J. Immunol. (1993) 23:2320-2325			
#title	A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.			
#cross-references	M01D:93380506			
#accession	I38005			
#status	preliminary; translated from GB/EMBL/DBJ			
#molecule_type	mRNA			
#residues	1-1156 #label RES			
CLASSIFICATION	#cros-ref EMBL:22970; NID:9312145; PID:9312146			
KEYWORDS	#superfamily scavenger receptor cysteine-rich domain			
FEATURE	cytosol			
43-147	#domain scavenger receptor cysteine-rich domain	homology		
151-254	#label SRC1\	#domain scavenger receptor cysteine-rich domain	homology	
258-361	#label SRC2\	#domain scavenger receptor cysteine-rich domain	homology	
365-468	#label SRC3\	#domain scavenger receptor cysteine-rich domain	homology	
470-573	#label SRC4\	#domain scavenger receptor cysteine-rich domain	homology	
575-678	#label SRC5\	#domain scavenger receptor cysteine-rich domain	homology	
711-814	#label SRC6\	#domain scavenger receptor cysteine-rich domain	homology	
816-920	#label SRC7\	#domain scavenger receptor cysteine-rich domain	homology	
921-1024	#label SRC8\	#domain scavenger receptor cysteine-rich domain	homology	
SUMMARY	#label SRC9	#domain scavenger receptor cysteine-rich domain	homology	
	#length 1156	#molecular-weight 125352	#checksum 9043	
Query Match	50.2%;	Score 380;	DB 2;	Length 1156;
Best Local Similarity 51.0%;	Pred. NO. 3,48e-65;			
Matches 51;	Conservative 19;	Mismatches 30;	Indels 0;	Gaps 0;
Db	714	LRVNGGRCAGRYEIIHESGWTICDDSDWLDSDAHVVCRLGCGEAINATGSAHFEGGT	773	
QY	447	VRLDGKKRGREYFINGOWGICDDGTBDKDAVIRQLGTYGPARFARTMAVFGSGK	506	
Db	774	GPWLDEKCKNGKESRIMWCHSHGQOQCRKEDAGVIC	813	
QY	507	GPVHDVVKCTGNERSLADCIKODIGRHNCRHSDDAGVIC	546	

RESULT 10
ENTRY A44407 #type complete
TITLE macrophage scavenger receptor, RSRL (collagen-like domain) -
ORGANISM rabbit
#formal_name Oryctolagus cuniculus #common_name domestic
#label format
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
16-Jul-1999
ACCESSIONS A44407
#authors DOI, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.;
Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.;
Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
J. Biol. Chem. (1993) 268:2126-2133
#journal Charged collagen structure mediates the recognition of
#file negatively charged macromolecules by macrophage scavenger
receptors.
#cross-references M01D:93131972
#accession A44407
#status preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-454 #label DOI
#experimental_source lung
#note sequence extracted from NCBI backbone (NCBIP:123206)
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor
cysteine-rich domain homology
FEATURE 350-453
#domain scavenger receptor cysteine-rich domain homology
#label SRC
SUMMARY #length 454 #molecular-weight 49735 #checksum 2435
Query Match 49.3%; Score 373; DB 2; Length 454;
Best Local Similarity 48.0%; Pred. No. 1,67e-63;
Matches 48; Conservative 20; Mismatches 32; Indels 0; Gaps 0;
DB 353 VRLVGRGPHGKREVEILANGOMGLVCDHMLRAGQVYCRSLGYKVSYHKKAYFGQGT 412
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 447 VRLMDGNNKRGREYEVINGQWGTICDDGWTDKDAVAICRLQGLGKPARATMAYFEGR 506.
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
DB 413 GPTWLNVEPCLGMSSIEECKIRQWGVYVCSHGSDAGVTC 452
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 507 GPIHVDNWKCTGNERSLADICIKDIGRHNCRHSEDAGVIC 546
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
#type complete
#formal_name Mus musculus #common_name house mouse
#label format
DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change
16-Jul-1999
ACCESSIONS A55840
#authors Elomaa, O.; Kangas, M.; Sahlberg, C.; Tuukkanen, J.;
Somunen, R.; Liakka, A.; Thesleff, I.; Kraal, G.;
Tryggvason, K.
Cell (1995) 80:603-609
#journal Cloning of a novel bacteria-binding receptor structurally
#file related to scavenger receptors and expressed in a subset of
macrophages.
#cross-references M01D:95171455
#accession A55840
#status preliminary
#molecule_type mRNA
#residues 1-518 #label ELO
#cross-references GB:U8424; NID:g682722; PID:g682723
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS transmembrane protein
FEATURE 418-518
#domain scavenger receptor cysteine-rich domain homology
#label SRC
SUMMARY #length 518 #molecular-weight 52730 #checksum 8159

Query Match 48.7%; Score 369; DB 2; Length 518;
Best Local Similarity 54.0%; Pred. No. 1.51e-62;
Matches 54; Conservative 17; Mismatches 24; Indels 5; Gaps 2;
DB 423 VRLMGTRNR--GRAEYVNNEMGTICDDMDNDATYFCMLGT---SRGRLSSYGGGS 477
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 447 VRLMDGNNKRGREYEVINGQWGTICDDGWTDKDAVAICRLQGLGKPARATMAYFEGR 506
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
DB 478 GNTMDNVNCRGTENSLMDCSKNSMGWNCVHNEDAGVEEC 517
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 507 GPIHVDNWKCTGNERSLADICIKDIGRHNCRHSEDAGVIC 546
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
#type complete
#formal_name Mus musculus #common_name house mouse
#label format
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
20-Sep-1999
ACCESSIONS B44407
#authors DOI, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.;
Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.;
Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
J. Biol. Chem. (1993) 268:2126-2133
#journal Charged collagen structure mediates the recognition of
#file negatively charged macromolecules by macrophage scavenger
receptors.
#cross-references M01D:93131972
#accession B44407
#status preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-458 #label DOI
#experimental_source macrophage-like cell line P38BD
#note sequence extracted from NCBI backbone (NCBIP:123207)
REFERENCE A38260
#authors Freeman, M.; Ashkenas, J.; Rees, D.J.G.; Kingsley, D.M.;
Copeland, N.G.; Jenkins, N.A.; Krieger, M.
Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8810-8814
#journal An ancient, highly conserved family of cysteine-rich protein
#file domains revealed by cloning type I and type II murine
macrophage scavenger receptors.
#cross-references M01D:91062370
#accession A38260
#status preliminary
#molecule_type mRNA
#residues 349-458 #label FRE
#cross-references GB:M59445; GB:M36817; NID:g192737; PIDN:AAA37464.1;
PID:g192738
REFERENCE I56334
#authors Ashkenas, J.; Penman, M.; Vasile, E.; Acton, S.; Freeman,
M.W.; Krieger, M.
J. Lipid Res. (1993) 34:983-1000
#journal Structures and high and low affinity ligand binding
#file properties of murine type I and type II macrophage
scavenger receptors.
#cross-references M01D:93359822
#accession I56334
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 5-458 #label RES
#cross-references GB:L04274; NID:g293745; PIDN:AAA39747.1; PID:g293746
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor
cysteine-rich domain homology
FEATURE 354-457
#domain scavenger receptor cysteine-rich domain homology
#label SRC
SUMMARY #length 458 #molecular-weight 50130 #checksum 1435
Query Match 48.6%; Score 368; DB 2; Length 458;
Best Local Similarity 49.0%; Pred. No. 2.63e-62;
Matches 49; Conservative 15; Mismatches 36; Indels 0; Gaps 0;

Db	ENTRY	RESULT	13
357	VLTVGSGAHEGVELEFHQOQCTICDDRDIDIRGAYCVSLGYQEVLDVAKRRHFGGCT		416
447	VALMDGKKEKRGVEVFINGQWCTICDDGWTDRDAVAICNOLGKGRARATMYAFGECK		506
417	GPIMLVNVCFGRESSIENCKINQMGVLSGCSHSDAGVTC		456
507	GPIMVNDVKTGNERSLADCIKODIGHNCRHSDAGVIC		546
DATE	ORGANISM		
18-Jun-1993	Bos primigenius taurus	#common_name	cattle
13-Nov-1998	sequence-revision	19-May-1994	#text_change
ACCESSIONS	REFERENCE		
A46496	A46496		
Wijnjaard, P.L.; Metzelaar, M.J.; Machugh, N.D.; Morrison, W.I.; Clevers, H.C.			
J. Immunol. (1992) 149:3273-3277			
Molecular characterization of the WC1 antigen expressed specifically on bovine CD4-CD8- gamma delta T lymphocytes.			
#cross-references	#accession		
WUID:93056489	A46496		
#status	preliminary		
molecule-type	mRNA		
#residues	1-1436	#label	W1J
#cross-references	EMBL:X63723; NID:913; PID:g14		
#experimental_source	CD4-CD8- gamma delta T lymphocytes		
#note	sequence extracted from NCBI backbone (NCBIP:117475)		
REFERENCE			
145834			
Wijnjaard, P.L.; Machugh, N.D.; Metzelaar, M.J.; Romberg, S.; Bensaïd, A.; Pepin, L.; Davis, W.C.; Clevers, H.C.			
J. Immunol. (1994) 152:3476-3482			
Members of the novel WC1 gene family are differentially expressed on subsets of bovine CD4-CD8- gamma delta T lymphocytes.			
#cross-references	WUID:94194107		
#accession	I45834		
#status	preliminary; translated from GB/EMBL/DBJ		
molecule-type	mRNA		
#residues	1-1436	#label	W12
#cross-references	EMBL:X63723; NID:913; PID:g14		
GENERIC			
gene	WC1.1		
CLASSIFICATION	#superfamily	scavenger receptor	cysteine-rich domain
FEATURE			
1-23	#domain	signal sequence	#status
24-1436	#product	WC1 antigen	#status
5-131	#domain	scavenger receptor	cysteine-rich domain
8-234	#domain	scavenger receptor	cysteine-rich domain
236-340	#domain	scavenger receptor	cysteine-rich domain
373-476	#domain	scavenger receptor	cysteine-rich domain
478-581	#domain	scavenger receptor	cysteine-rich domain
583-686	#domain	scavenger receptor	cysteine-rich domain
687-789	#domain	scavenger receptor	cysteine-rich domain
791-895	#domain	scavenger receptor	cysteine-rich domain
928-1031	#domain	scavenger receptor	cysteine-rich domain
1033-1136	#domain	scavenger receptor	cysteine-rich domain
1152-1255	#domain	scavenger receptor	cysteine-rich domain
SUMMARY			
#length	1436	#molecular-weight	154196
		#checksum	7163

Query Match	47.8%	Score 362	DB 2	Length 1436
Best Local Similarity	53.0%	Prod. No. 7.15e-61		
Matches	53	Conservative	14	Mismatches 33; Indels 0; Gaps 0;
Db	376	LRVYDGGPCAGRVEILIDOGSMGICDDGWDLDARVRCROLCGCEALNATGSAHFAGS	435	
Oy	447	VRLDMDGENKKEGRVEYFINGOMGICDDGTDKDAVICROLGYPAPARPMAYFEGK	506	
Db	436	GPIMLDNINCTGKESHVRCPSRGWGCHCRKAKODAGVIC	475	
Oy	507	GPIMVDNKCCTGNERSLADICIKODIGHNCRHSEDAGYIC	546	
RESULT	14			
ENTRY	A38415	#type complete		
TITLE	macrophage scavenger receptor splice form I - human			
ORGANISM	#formal_name Homo sapiens #common_name man			
DATE	28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 20-Sep-1999			
ACCESSIONS	A38415			
REFERENCES	A38415			
authors	Matsumoto, A.; Naito, M.; Itakura, H.; Ikemoto, S.; Asaoka, H.; Hayakawa, I.; Kanamori, H.; Aburatani, H.; Takaku, F.; Suzuki, H.; Kobari, Y.; Miyai, T.; Takahashi, K.; Cohen, E.H.; Wydro, R.; Housman, D.E.; Kodama, T.; Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9133-9137			
#journal	Human macrophage scavenger receptors: primary structure, expression, and localization in atherosclerotic lesions.			
#title	expression, and localization in atherosclerotic lesions.			
#cross-references	MUTID:91067661			
#accession	A38415			
#status	preliminary			
#molecule_type	mRNA			
#residues	1-451 ##label MAT			
#cross-references	GB:D90187; NID:g219989; PIDN:BAJ4208.1; PID:01014913; PID:g219990			
REFERENCE	A44408			
authors	Emi, M.; Asaoka, H.; Matsumoto, A.; Itakura, H.; Kurihara, Y.; Wado, Y.; Kanamori, H.; Yezaki, Y.; Takahashi, E.; Lepetit, M.; Lalouel, J.; Kodama, T.; Mukai, T.			
#journal	J. Biol. Chem. (1993) 268:2120-2125			
#title	Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene.			
#cross-references	MUTID:93131971			
#accession	A44408			
#status	preliminary: not compared with conceptual translation			
#molecule_type	nucleic acid			
#residues	155-872 ##label EMI			
#note	sequence extracted from NCBI backbone (NCBIP:123189)			
GENETICS				
gene	GDB:MSR1			
#cross-references	GDB:128046; OMIM:153622			
CLASSIFICATION	#map_position 8p22-8p22			
	#superfamily unassigned collagens; scavenger receptor cysteine-rich domain homology			
KEYWORDS	alternative splicing; coiled coil; transmembrane protein			
FEATURE				
347-450	#domain scavenger receptor cysteine-rich domain homology			
	#label SRC			
SUMMARY	#length 451 #molecular-weight 49762 #checksum 6168			
Query Match	47.7%	Score 361	DB 2	Length 451
Best Local Similarity	49.0%	Prod. No. 1.24e-60		
Matches	49	Conservative	17	Mismatches 34; Indels 0; Gaps 0;
Db	350	VRVYGGSGPHEGRVEILHSGOWGTCIDDDREWEVRGVGVCRSLGYPGVAVHKAHFGGCT	409	
Oy	447	VRLDMDGENKKEGRVEYFINGOMGICDDGTDKDAVICROLGYPAPARPMAYFEGK	506	
Db	410	GPIMLDNINCTGKESHVRCPSRGWGCHCRKAKODAGVIC	449	
Oy	507	GPIMVDNKCCTGNERSLADICIKODIGHNCRHSEDAGYIC	546	

```

RESULT      15
ENTRY       A53202 #type complete
TITLE       cyclophilin C-associated protein MAMA/CyCAP precursor - mouse
ALTERNATE_NAMES MAC-2-binding glycoprotein homolog; murine adherent
               macrophage (MAMA) protein
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change
               16-Jul-1999
ACCESSIONS  A53202; A48231; S33136
REFERENCE   A53202
#authors    Chicheportiche, Y.; Vassalli, P.
#journal    J. Biol. Chem. (1994) 269:5512-5517
#title      Cloning and expression of a mouse macrophage cDNA coding for
               a membrane glycoprotein of the scavenger receptor
               cysteine-rich domain family.
#cross-references MIMD:94164889
#accession    A53202
               ##molecule_type mRNA
               ##residues 1-577 ##label CHI
               ##cross-references EMBL:X67809; NID:9297032; PID:9297033; GB:X67803
               #note      sequence appears consistent with either a GPI anchor or
               ##note      transmembrane domain near the carboxyl end
REFERENCE   A48231
#authors    Friedman, J.; Trahey, M.; Weissman, T.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6815-6819
#title      Cloning and characterization of cyclophilin C-associated
               protein: a candidate natural cellular ligand for
               cyclophilin C
#cross-references MIMD:93342080
#accession    A48231
               ##molecule_type mRNA
               ##residues 1-24, 'G', '26-227', 'P', '229-465', 'NE', '468-572', 'LH' ##label, ~
               #note      FRI
               ##cross-references GB:LI6894; NID:9397799; PID:9397800
GENETICS    #cross-references GB:LI6894; NID:9397799; PID:9397800
CLASSIFICATION #gene      CyCAP
               #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS     cell surface component; glycoprotein; membrane protein
FEATURE      1-18      #domain signal sequence #status predicted #label SIG\
21-124         #domain scavenger receptor cysteine-rich domain homology
               #label SRC
SUMMARY      #length 577 #molecular-weight 64491 #checksum 6506
ery Match    47.2%; Score 357; DB 2; Length 577;
st Local Similarity 50.0%; Pred. No. 1.12e-59;
tches 50; Conservative 16; Mismatches 34; Indels 0; Gaps 0;
Db 24 MRLVNGASANGRAVEIFYRGRMGTVCDNLMLDAHVVCRAALGYENATQALGRAAFGPGK 83
:||||: : |||||:| :|||:|:| : |||||:| :|||:|:| :|||:|:| :|||:|
Oy 447 VRLMDGENKKRGREVEFINQWGTICDDGWTKDAVAICRQLGKPARARTMAYFGEGR 506
|||:|:| :|||:|:| :|||:|:| :|||:|:| :|||:|:| :|||:|:| :|||:|
Db 84 GPIMLDEVECTGTESSLASCSRLGMWVSRGCGHEKDAGYVC 123
|||:|:| :|||:|:| :|||:|:| :|||:|:| :|||:|:| :|||:|:| :|||:|
507 GPIMVNVKCTIGNERSLADCIKODIGRHNCRHSEADAGYIC 546

```

Search completed: Mon Mar 13 10:34:39 2000
Job time : 12 secs.

THIS PAGE BLANK (USPTO)

SCRCY
-947-6-06.rsp
1/17/03 4:03 PM

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Mon Mar 13 10:33:52 2000; MasPar time 16.67 Seconds
Tabular output not generated.

```

Title: >US-09-147-947-6
Description: (447-547) from US09147947A.pep (6 of 6)
Perfect Score: 757
Sequence: 1 VRLMDGKKEKGRPEVING.....KODIGRHNCRHSDEAGVTCID 101

```

Scoring table: PAM 150

Searched: 225878 seqs, 69334122 residues

Listing first 45 summaries

Database: **spremb112**

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_Ornithelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 37.831; Variance 59.275; scale 0.638

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	466	61.6	2153	5	097375	SCAVENGER RECEPTOR CYS	7.20e-90
2	411	54.3	1036	5	097378	SCAVENGER RECEPTOR CYS	6.99e-76
3	404	53.4	2043	5	096949	SRCR DOMAIN, MEMBRANE	4.10e-74
4	400	52.8	531	5	017064	SRP85.	4.19e-73
5	391	51.7	585	4	008380	MAC-2 BINDING PREPROTEIN	7.74e-71
6	380	50.2	1116	4	007898	M130 ANTIGEN PRECURSOR	4.49e-68
7	380	50.2	1149	4	007901	M130 ANTIGEN, EXTRACEL	4.49e-68
8	380	50.2	1151	4	007899	M130 ANTIGEN, CYTOPLAS	4.49e-68
9	380	50.2	1156	4	007900	M130 ANTIGEN, CYTOPLAS	4.49e-68
10	374	49.4	578	11	P70117	PANCREAS CANCER-ASSOCI	1.43e-66
11	372	49.1	574	11	070513	MAMA.	4.34e-66
12	369	48.7	518	11	060754	BACTERIA BINDING MACRO	2.56e-65
13	369	48.7	2083	11	060979	CRP-DUCIN PRECURSOR (2.56e-65
14	368	48.6	734	4	095Y8	LYSYL OXIDASE-LIKE PRO	4.54e-65
15	368	48.6	774	4	094XK0	LYSYL OXIDASE-RELATED	4.54e-65
16	366	48.3	1785	4	09Y4V9	DMBT1/56B.1 PROTEIN PR	1.44e-64
17	366	48.3	1785	4	09Y211	DMBT1 PROTEIN.	1.44e-64
18	365	48.2	356	6	097682	UNKNOWN MRNA, PARTIAL	2.55e-64
19	364	48.1	822	13	09YHCl	LYSYL OXIDASE HOMOLOG	4.34e-64
20	364	48.1	895	13	09Y6M1	LYSYL OXIDASE RELATED	4.34e-64

1	357	47.2	577	11	0075757	PEPTIDYLPROLYL ISOMERASE	2.53E-65
2	356	47.0	528	5	0973739	SCAVENGER RECEPTOR CY5	4.49E-65
21	353	46.6	918	13	0920988	PEMA-SRCR PROTEIN PREC	2.51E-67
22	352	46.5	1594	6	0952181	HENSIN.	4.44E-61
23	351	46.4	532	5	0251111	A SPERM-ACTIVATING PEP	7.88E-61
24	350	46.2	546	6	0291111	SCAVENGER-RECEPTOR PRO	1.40E-66
25	350	46.2	1290	11	0628277	ENBERIN.	1.40E-66
26	350	46.2	574	11	0356495	CYCLOPHILIN C-ASSOCIAT	1.38E-58
27	346	45.7	600	6	0289110	MODIN (FRAGMENT).	2.44E-59
28	345	45.6	504	6	0289088	MODIN (FRAGMENT).	2.44E-59
29	343	45.3	369	6	0291110	SCAVENGER-RECEPTOR PRO	2.39E-58
30	341	45.0	462	6	0288811	MEMBRANE PROTEIN SCAVE	4.24E-58
31	340	44.9	754	11	0921755	LYSYL OXIDASE-RELATED	4.24E-58
32	340	44.9	483	11	0920829	MACROPHAGE RECEPTOR MA	2.29E-58
33	340	44.9	347	4	0386666	SP ALPHA.	4.04E-56
34	333	43.9	127	6	0953167	M130 ANTIGEN (FRAGMENT	6.93E-55
35	332	43.9	520	4	0955533	MACROPHAGE RECEPTOR.	1.22E-54
36	332	43.1	592	4	09Y4K77	CP6E.	6.70E-54
37	333	42.7	595	4	09Y4K99	CP6C.	6.70E-54
38	333	42.7	601	4	09Y4K88	CP6D.	6.70E-54
39	333	42.7	635	4	09Y4L00	CP6B.	6.70E-54
40	323	42.7	352	11	0353011	SP-ALPHA.	3.53E-55
41	316	41.7	352	11	0353003	SP-ALPHA.	3.53E-55
42	316	41.7	665	11	0610003	T-CELL DIFFERENTIATION	3.52E-55
43	316	41.7	804	6	0291113	SCAVENGER-RECEPTOR PRO	3.22E-50
44	308	40.7					

ALIGNMENTS

ID	RESULT	PRELIMINARY;	PRT;	2153 AA.
ID	097375			
AC	097375:			
DT	01-MAY-1999 (TREMBLrel. 10. Created)			
DT	01-MAY-1998 (TREMBLrel. 10. Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12. Last annotation update)			
DE	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.			
GN	SRCL12.			
OS	Strongylocentrotus purpuratus (Purple sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;			
OC	Echinozoidea; Echinacea; Echinoidea; Strongylocentrotidae;			
OC	Strongylocentrotus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	PANCER Z., RAST J., DAVIDSON E.H.;			
RT	"Evolution of the Immune System: Transcription Factors and Downstream			
RT	Genes of the Mammalian Immune System Expressed in Sea Urchin			
RT	Ceolomocytes.";			
RL	Immunogenetics 0:0-0(1999).			
DR	EMBL; AF064259; AAD08547.			
DR	PROSITE; PS00420; SPERACT_RECEPTOR; 15.			
KW	Signal; Receptor.			
FT	SIGNAL	1	16	POTENTIAL.
FT	CHAIN	17	2153	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN
FT				TYPE 12.
SQ	SEQUENCE	2153 AA;	226557 MM;	CAB3A8A9 CRC32;
	Query Match	61.6%;	Score 466;	DB 5; Length 2153;
	Best Local Similarity	60.0%;	Pred. No. 7.20e-90;	
	Matches	50; Conservative	13; Mismatches	27; Indels 0; Gaps 0
Db	1837	VRLFGGNNHNGRGVFLAGTWGTCDDGWDIDASVYCRHLGYSASRATNATFGGGE	1896	
Oy	447	VRLMDGKNKKERAEVFLINGWGTCDDGWDKDAVAICRQIGYGRAPARTMAVFGGK	506	
Db	1897	GIIVMDGVSCDGDDETDLSCMHAGIGIFHNGCHOEDAGVVC	1936	
Oy	507	GPIHDVNYKCTGNERSLADCIKODIGIRANCRHSDEAGVYC	546	
RESULT	2			
ID	097378	PRELIMINARY;	PRT;	1036 AA.
AC	097378:			
DT	01-MAY-1999 (TREMBLrel. 10. Created)			

01-MAY-1999 (Tremblrel. 10, last sequence update)
 01-NOV-1999 (Tremblrel. 12, last annotation update)
 SCVENNER RECEPTOR CYSTEINE-RICH PROTEIN PRECURSOR.
 Strongylocentrotus purpuratus (Purple sea urchin).
 Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 Strongylocentrotus.
 [1]
 SEQUENCE FROM N.A.
 RA PANCER 2., RAST J.P., DAVIDSON E.H.;
 "Origins of Immunity: Transcription Factors and Effector Genes of the
 Vertebrate Immune System Expressed in Sea Urchin Coelomocytes.";
 Immunogenetics 0:0-0(1999).
 DR EMBL, AF076513; AAD05493.1; -.
 KW Signal; Receptor.
 FT SIGNAL
 CHAIN 1 20
 SEQUENCE 1036 AA; 109704 MW; CBE8C531 CRC32;
 SCVENNER RECEPTOR CYSTEINE-RICH PROTEIN.
 Query Match 54.3%; Score 411; DB 5; Length 1036;
 Best Local Similarity 54.0%; Pred. No. 6,99e-76;
 Matches 54; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

01-MAY-1999 (Tremblrel. 10, last sequence update)
 01-NOV-1999 (Tremblrel. 10, last sequence update)
 01-NOV-1999 (Tremblrel. 12, last annotation update)
 SRCR DOMAIN, MEMBRANE FORM 2.
 GN SRCRM2.
 OS Geodia cydonium (Sponge)
 CC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
 CC Astrophorida; Geodidae; Geodia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98369060.
 RA BLUMBACH B., PANCER Z., DIEHL-SEIFERT B., STEFFEN R., MUEKNER J.,
 MUELLER I., MUELLER W.E.G.;
 "The putative sponge aggregation receptor. Isolation and
 characterization of a molecule composed of scavenger receptor
 cysteine-rich domains and short consensus repeats.";
 J. Cell Sci. 111:2635-2644(1998).
 DR EMBL, Y14853; CAA75175.1; -.
 DR HSSP; P10998; IVC.
 DR PROSITE; PS00420; SPERACT-RECEPTOR; 8.
 SQ SEQUENCE 2043 AA; 220896 MW; 4CE19401 CRC32;

Query Match 53.4%; Score 404; DB 5; Length 2043;
 Best Local Similarity 54.5%; Pred. No. 4.10e-74;
 Matches 55; Conservative 16; Mismatches 30; Indels 0; Gaps 0;

RESULT 4 PRELIMINARY: PRT; 531 AA.

01-NOV-1996 (Tremblrel. 01, Created)
 01-NOV-1996 (Tremblrel. 01, last sequence update)
 01-NOV-1999 (Tremblrel. 12, last annotation update)
 SP85.
 OS Arabidopsis punctulata (Punctate sea urchin).
 CC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 CC Euechinoidea; Echinacea; Artabacoida; Artabacidae; Artabac.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA DANGOTT L.J., RANKIN T.L.;
 Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89184581.
 RA DANGOTT L.J., JORDAN J.E., BELLET R.A., GARBERS D.L.;
 "Cloning of the mRNA for the protein that crosslinks to the egg
 peptide speract.";
 Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).
 DR EMBL; U30933; AAA7510.1; -.
 DR PFAM; PF00530; SRCR; 4.
 DR PRINTS; PR00258; SPERACTRCPTR.
 SQ SEQUENCE 531 AA; 57414 MW; 0FE36E2D CRC32;

Query Match 52.8%; Score 400; DB 5; Length 531;
 Best Local Similarity 56.3%; Pred. No. 4.19e-73;
 Matches 58; Conservative 13; Mismatches 28; Indels 4; Gaps 4;

01-NOV-1996 (Tremblrel. 01, Created)
 01-NOV-1996 (Tremblrel. 01, last sequence update)
 01-NOV-1999 (Tremblrel. 12, last annotation update)
 MAC-2 BINDING PROTEIN PRECURSOR.
 DE Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93300818.
 RA KOTHS K., TAYLOR E., HALENBECK R., CASIPIT C., WANG A.;
 "Cloning and characterization of a human Mac-2-binding protein, a new
 member of the superfamily defined by the macrophage scavenger receptor
 cysteine-rich domain.";
 J. Biol. Chem. 268:14245-14249(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94308070.
 RA ULIRICH A., SURES I., D'EGIDO M., JALLAL B., POWELL T.J., HERBST R.,
 DREBS A., AZAM M., ROBINSTEIN M., NATOLI C.;
 "The secreted tumor-associated antigen 90K is a potent immune
 stimulator.";
 J. Biol. Chem. 269:18401-18407(1994).
 DR EMBL; L13210; AAA36193.1; -.
 DR EMBL; X79089; CAA55699.1; -.
 DR PFAM; PF00530; SRCR; 1.
 DR PRINTS; PR00258; SPERACTRCPTR.
 KW Signal.
 FT SIGNAL
 CHAIN 1 18
 SEQUENCE 585 AA; 65330 MW; AAE9E32 CRC32;

Query Match 51.7%; Score 391; DB 4; Length 585;


```

RESULT 13
ID Q60997 PRELIMINARY: PRT: 2083 AA.
AC Q60997:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE CRP-DUCTIN PRECURSOR (CRP).
GN CRP OR CRP OR CRP-DUCTIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
SEQUENCE FROM N.A.
STRAIN-BALB/C; TISSUE=JEJUNUM;
MEDLINE: 96362470.
CHENG H., BJERKNES M., CHEN H.;
"CRP-ductin: a gene expressed in intestinal crypts and in pancreatic
and hepatic ducts.";
RL Anat. Rec. 244:327-343(1996).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN (CRP-ALPHA AND CRP-
BETA) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE
SEQUENCE SHOWN HERE IS THAT OF CRP-ALPHA.
CC -1- TISSUE SPECIFICITY: INTESTINE AND PANCREAS. IN THE COLON,
EXPRESSED IN CRYPT CELLS BUT NOT MATURE VILLOSUS CELLS. MOST HIGHLY
EXPRESSED IN MID-CRYPT. ALSO EXPRESSED IN EPITHELIUM LINING
HEPATIC AND PANCREATIC DUCTS.
CC EMBL: U37448; AAC52505.1; -.
DR HSP: P29392; ISEP.
DR MGD: MGI:106210; Crpd.
DR PFAM: PF00431; CUB; 5.
DR PFAM: PF00530; SRCR; 8.
DR PFAM: PF00100; zona_pellucida; 1.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Signal; Transmembrane; Alternative splicing.
FT SIGNAL 1 28
FT CHAIN 29 2083
FT DOMAIN 29 2042 EXTRACELLULAR.
FT DOMAIN 986 1017 THIR-RICH.
FT DOMAIN 1250 1259 PRO/THR-RICH.
FT DOMAIN 1610 1625 PRO/SER/THR-RICH.
FT TRANSMEM 2043 2065 POTENTIAL.
FT DOMAIN 2066 2083 CYTOPLASMIC.
FT VARSPLIC 2030 2083 LAVEDKKPSSAVYPTAIRIGVFLAMVLVAFTIGRR
THIDGQRPSTKL -> PELSSQSGPSSH (IN
CRP-BETA).
SEQUENCE 2083 AA; 226734 MW; 945FC070 CRC32;
Query Match 48.7%; Score 369; DB 11; Length 2083;
Best Local Similarity 52.0%; Pred. No. 2,56e-65;
Matches 52; Conservative 16; Mismatches 32; Indels 0; Gaps 0;
DB 1508 LRLVNGSNCEGRVVELLYGSMGTVCDDSGISDANVVCROGGSALSPGNAMFGGS 1567
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 447 VRLMDENKKEGRVEYFINGWGITICDDGWTDKDAVICRQGYGKPARAFMAVFGEK 506
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1568 GLIVDDVSCSGYESTHLNMGCHPGLVHNCRHSDEAGVIC 1607
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 507 GPRIHDNVKCTGNERSLADCIKODIGRHNCRHSEDAVIC 546
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 14
ID Q9Y5X8 PRELIMINARY: PRT: 638 AA.
AC Q9Y5X8:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE LYSYL OXIDASE-LIKE PROTEIN 2.
GN LOXL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN (1)
SEQUENCE FROM N.A.
RP TISSUE=SPLEEN, PLACENTA;
RC TISSUE=SPLEEN, PLACENTA;
RX MEDLINE: 99230328.
RA SAUV C.J., TRONECKER H., BOGIC L., BRYANT-GREENWOOD G.D., BOYD C.D.,
CSISZAR K.;
"the LOXL2 gene encodes a new lysyl oxidase-like protein and is
expressed at high levels in reproductive tissues.";
RL J. Biol. Chem. 274:12939-12944(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN, PLACENTA;
RA JOORDAN-LE SAUV C., CSISZAR K.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF117949; AAD34343.1; -.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
SQ SEQUENCE 638 AA; 71101 MW; C232E6N4 CRC32;
Query Match 48.6%; Score 368; DB 4; Length 638;
Best Local Similarity 51.0%; Pred. No. 4,54e-65;
Matches 52; Conservative 21; Mismatches 26; Indels 3; Gaps 3;
DB 125 VALRGCAVIGREYVLKNGEYTCDDKMDLVSAVYCRELGF-GSAKEAVTGSRLGOG 183
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 447 VRLMDENKKEGRVEYFINGWGITICDDGWTDKDAVICRQGYGKPAR-ARTMAVFGEK 505
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 184 IGPILHNEIOCTGNEKSIIDC-KFNAESOGCNHEEDAGRCN 224
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 506 KGPRIHDNVKCTGNERSLADCIKODIGRHNCRHSEDAVIC 547
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 15
ID Q9Y4K0 PRELIMINARY: PRT: 774 AA.
AC Q9Y4K0:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE LYSYL OXIDASE-RELATED PROTEIN.
GN MS9-14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
SEQUENCE FROM N.A.
RA SATO H., PAPACONSTANTINOU J., SATO H., GOLDSTEIN S.;
"Regulation of a Novel Gene Encoding a Lysyl Oxidase-Related Protein
in Cellular Adhesion and Senescence.";
RL J. Biol. Chem. 0:0-0(1997).
DR EMBL: U89942; AAB49697.1; -.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
SQ SEQUENCE 774 AA; 86724 MW; 1A5C47AA CRC32;
Query Match 48.6%; Score 368; DB 4; Length 774;
Best Local Similarity 51.0%; Pred. No. 4,54e-65;
Matches 52; Conservative 21; Mismatches 26; Indels 3; Gaps 3;
DB 326 VALRGCAVIGREYVLKNGEYTCDDKMDLVSAVYCRELGF-GSAKEAVTGSRLGOG 384
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 447 VRLMDENKKEGRVEYFINGWGITICDDGWTDKDAVICRQGYGKPAR-ARTMAVFGEK 505
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 385 IGPILHNEIOCTGNEKSIIDC-KFNAESOGCNHEEDAGRCN 425
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 506 KGPRIHDNVKCTGNERSLADCIKODIGRHNCRHSEDAVIC 547
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: Mon Mar 13 10:34:10 2000
Job time : 18 secs.

THIS PAGE BLANK (USPTO)

 RELEASE

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Search - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 13 10:33:26 2000; Maspar time 6.50 Seconds
 Tabular output not generated. 464,390 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (447-547) from US09147947A.pep (6 of 6)
 Perfect Score: 757
 Sequence: 1 VALMDGKMKKEGVEVING.....KQDIGHKCHSEAGVIGD 101

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 38.951; Variance 57.048; scale 0.683

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Hit No.	Score	Query Match	Length	ID	Description	Pred. No.
1	757	100.0	875	1	NETR_HUMAN NEUROTRYP SIN PRECURSOR	7.67e-117
2	724	95.6	761	1	NETR_HUMAN NEUROTRYP SIN PRECURSOR	1.50e-167
3	385	50.9	453	1	MSRE_BOVIN MACROPHAGE SCAVENGER R	5.00e-74
4	381	50.3	454	1	MSRE_RABIT MACROPHAGE SCAVENGER R	5.89e-73
5	368	48.6	458	1	MSRE_MOUSE MACROPHAGE SCAVENGER R	1.74e-69
6	362	47.8	1436	1	WC11_BOVIN ANTIGEN WC1.1	6.88e-68
7	361	47.7	451	1	MSRE_HUMAN MACROPHAGE SCAVENGER R	1.27e-67
8	351	46.4	532	1	SPER_STRU EGG PEPTIDE SPERACT RE	5.73e-65
9	323	42.7	468	1	CD6_HUMAN T-CELL DIFFERENTIATION	1.38e-57
10	157	20.7	1034	1	ENTK_PIG ENTEROPEPTIDASE PRECUR	2.57e-16
11	146	19.3	583	1	CFAI_HUMAN COMPLEMENT FACTOR I PR	7.47e-14
12	136	18.0	1069	1	ENTK_MOUSE ENTEROPEPTIDASE (EC 3.	1.15e-11
13	125	16.5	1035	1	ENTK_BOVIN ENTEROPEPTIDASE PRECUR	2.51e-09
14	120	15.9	1019	1	ENTK_HUMAN ENTEROPEPTIDASE PRECUR	2.73e-08
15	113	14.9	495	1	CD5_BOVIN T-CELL SURFACE GLYCOP	7.16e-07
16	106	14.0	492	1	TMS2_HUMAN TRANSMEMBRANE PROTEASE	1.71e-05
17	96	12.7	495	1	CD5_HUMAN T-CELL SURFACE GLYCOP	2.01e-03
18	95	12.5	494	1	CD5_MOUSE T-CELL SURFACE GLYCOP	1.06e-02
19	91	12.0	274	1	BPAL_MOUSE NON-HAEM BROMOPEROXIDA	1.59e-02
20	90	11.9	378	1	VR2_HP130 REGULATORY PROTEIN E2.	7.84e-02
21	86	11.4	404	1	6P12_LYCES STYLE DEVELOPMENT-SPEC	1.71e-01
22	84	11.1	1342	1	XDH_DROPS XANTHINE DEHYDROGENASE	1.71e-01
23	84	11.1	1344	1	XDH_DROSO XANTHINE DEHYDROGENASE	1.71e-01

24	83	11.0	650	1	P72_HUMAN PROBABLE RNA-DEPENDENT	2.52e-01
25	81	10.7	287	1	NIFH_FRAP NITROGENASE IRON PROTE	5.38e-01
26	81	10.7	305	1	MIAA_AOUAE TRNA DELTA(2)-ISOPENTE	5.38e-01
27	81	10.7	1335	1	XDH_DROME XANTHINE DEHYDROGENASE	5.38e-01
28	81	10.7	1758	1	YXR7_YEAST HYPOTHEICAL 197.5 KD	5.38e-01
29	81	10.7	1758	1	YJW5_YEAST HYPOTHEICAL 197.6 KD	5.38e-01
30	80	10.6	380	1	KAP1_MOUSE CAMP-DEPENDENT PROTEIN	7.82e-01
31	80	10.6	380	1	KAP1_MOUSE CAMP-DEPENDENT PROTEIN	7.82e-01
32	80	10.6	380	1	KAP1_MOUSE CAMP-DEPENDENT PROTEIN	7.82e-01
33	80	10.6	514	1	ABF2_BACOV ALPHA-L-ARABINOFURANOS	7.82e-01
34	79	10.4	507	1	DAF_CAVPO COMPLEMENT DECAT-ACCEL	1.13e+00
35	78	10.3	287	1	NIFH_FRAP NITROGENASE IRON PROTE	1.64e+00
36	78	10.3	514	1	IMD2_MOUSE INOSINE-5'-MONOPHOSPA	1.64e+00
37	77	10.2	453	1	YKP9_KILDA HYDROXYMETHYLGLUTARYL-	2.35e+00
38	77	10.2	491	1	HMCS_YEAST T-CELL SURFACE GLYCOP	2.35e+00
39	77	10.2	491	1	CD5_RAT T-CELL SURFACE GLYCOP	2.35e+00
40	77	10.2	697	1	YHFO_YEAST HYPOTHEICAL 79.0 KD P	2.35e+00
41	77	10.2	1162	1	VGL2_IYVA E2 GLYCOPROTEIN PRECUR	2.35e+00
42	77	10.2	1363	1	XDH_EMENT XANTHINE DEHYDROGENASE	2.35e+00
43	76	10.0	192	1	RPCL_BP186 REPRESSOR PROTEIN CI.	3.37e+00
44	76	10.0	383	1	METL_YEAST S-ADENOSYLMETHIONINE S	3.37e+00
45	75	9.9	550	1	MICO_STRCT MICOLISIN PRECURSOR (E	4.81e+00

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	875 AA.
ID	NETR_HUMAN			
AC	P67330;			
DT	15-DEC-1999 (Rel. 39, Created)			
DT	15-DEC-1999 (Rel. 39, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	NEUROTRYP SIN PRECURSOR (EC 3.4.21.-) (MOTOPROIN).			
GN	PSS312.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
NC	Eutheria; Primates; Catarrhini; Homidae; Homo.			
RC	[1]			
RP	TISSUE-BRAIN: N.A.			
RX	MEDLINE: 98201705.			
RA	PROBA K., GSCHEWEND T.P., SONDERGGER P.,			
RT	"Cloning and sequencing of the cDNA encoding human neurotysin.",			
RL	Biochim. Biophys. Acta 1396:143-147(1998).			
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC			
CC	ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH			
CC	LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.			
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.			
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL: AJ001531; CA004816.1; -			
DR	PROSITE: PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE: PS00135; TRYPSIN_SER; 1.			
DR	PROSITE: PS00420; SPERACT_RECEPTOR; 3.			
KW	Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.			
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	875	NEUROTRYP SIN.
FT	DOMAIN	23	92	PROLINE-RICH.
FT	DOMAIN	93	165	KRINGLE.
FT	DOMAIN	170	271	SRCR 1.
FT	DOMAIN	280	381	SRCR 2.
FT	DOMAIN	387	487	SRCR 3.

SCR4
 Adv. No. 3 is available

```

FT DOMAIN 500 601 SRCR 4
FT DOMAIN 619 875 SERINE PROTEASE.
FT DOMAIN 619 875 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 67D5272B CRC32:

Query Match 100.0%; Score 757; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 7,57e-177;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 500 VRLMDGNKKEGVEVFNQMGWTCDDGWTDKDAAVICRQLYGKPARARTMAFGEGK 559
447 VRLMDGNKKEGVEVFNQMGWTCDDGWTDKDAAVICRQLYGKPARARTMAFGEGK 506
560 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYCD 600
507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYCD 547

RESULT 2 STANDARD; PRT; 761 AA.
ID NETR_MOUSE 008762;
AC 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN) (BRAIN-SPECIFIC SERINE
PROTEASE 3) (BSSP-3).
OS PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GSCHEWEND T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neutrotypin, a novel multidomain serine protease expressed in the
nervous system.";
RL Mol. Cell. Neurosci. 9:207-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSURUOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
kringle-like structure and three scavenger receptor cysteine-rich
motifs.";
RL Biochem. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
AND AMYGDALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: Y13192; CAJ73646.1; -
DR EMBL: D89871; BAA23986.1; -
DR MGD; MGI:1100861; PRSS12.
DR PFM; PFO0530; SRCR; 3.
DR PFM; PFO0089; trypsin; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.
KW Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 761 NEUTROTYPIN.
FT DOMAIN 85 157 KRINGLE.
FT DOMAIN 166 267 SRCR 1.
FT DOMAIN 273 373 SRCR 2.
FT DOMAIN 386 487 SRCR 3.
FT DOMAIN 505 761 SERINE PROTEASE.
FT DOMAIN 505 516 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 516 517 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 562 562 CHARGE RELAY SYSTEM.
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 POTENTIAL.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA: 84118 MW: 3F3C4F35 CRC32:

Query Match 95.6%; Score 724; DB 1; Length 761;
Best Local Similarity 92.1%; Pred. No. 1,50e-167;
Matches 93; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 386 IRLVDGNKKEGVEVFNQMGWTCDDGWTDKDAAVICRQLYGKPARARTMAFGEGK 445
447 VRLMDGNKKEGVEVFNQMGWTCDDGWTDKDAAVICRQLYGKPARARTMAFGEGK 506
446 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYCD 486
507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYCD 547

RESULT 3 STANDARD; PRT; 453 AA.
ID MSRE_BOVIN
AC P21758;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
IDL RECEPTOR I AND II).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-LUNG;
RX MEDLINE: 90136965.
RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
RA KRIEGER M.;
RT "Type I macrophage scavenger receptor contains alpha-helical and
collagen-like coiled coils.";
RL Nature 343:531-535(1990).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE-LUNG;
RX MEDLINE: 90136973.
RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIEGER M.;
RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
scavenger receptor type II.";
RL Nature 343:570-572(1990).
CC -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROSCLEROSIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING

```

CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X51689; CAA35987.1; -
 CC EMBL: X54183; CAA38108.1; -
 CC PIR: S08276; S08276.
 CC PIR: S08278; S08278.
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC DR PFAM: PF00530; SRCR; 1.
 CC DR PFAM: PF01391; Collagen; 1.
 CC KM Transmembrane; Glycoprotein; 1.
 CC FT DOMAIN 1 50
 CC FT TRANSMEM 1 50
 CC FT DOMAIN 77 76
 CC FT DOMAIN 77 453
 CC FT DOMAIN 109 271
 CC FT DOMAIN 272 343
 CC FT DOMAIN 344 453
 CC FT CARBOHYD 82 82
 CC FT CARBOHYD 101 101
 CC FT CARBOHYD 142 142
 CC FT CARBOHYD 183 183
 CC FT CARBOHYD 220 220
 CC FT CARBOHYD 248 248
 CC FT CARBOHYD 266 266
 CC FT VARSPLIC 348 349
 CC FT VARSPLIC 350 453
 CC SQ SEQUENCE 453 AA; 50056 MW; 2CDE1EE1 CRC32;
 CC -----
 CC Query Match 50.98; Score 385; DB 1; Length 453;
 CC Best Local Similarity 50.08; Pred. No. 5,00e-74;
 CC Matches 50; Conservative 17; Mismatches 33; Indels 0; Gaps 0;
 CC 352 VRLVGGSGPHEGVEIFHEGOWGTVCDDRWELRGLVCRSLGYKQGVSHKRAYGKGT 411
 CC 447 VRLMDENKKEGVEVFINGOWGTICDDGWTKDAVAIVRQLGYKGPAPARIMAYFGEGK 506
 CC DB 412 GPIWNEVFCFKESSIEECRIKRWGVYACSHSDAGVTC 451
 CC 507 GPIHVDNVKCTGNSRLADCIKODIGRHNCRHSDAGVIC 546
 CC -----
 CC RESULT 4
 CC ID MSRE_RABIT STANDARD; PRT; 454 AA.
 CC AC 005585;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
 CC DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 CC DE LDL RECEPTOR I AND II).
 CC GN MSRL.
 CC OS Oryctolagus cuniculus (Rabbit).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 9301687.
 CC RA BICKEL P.E., FREEMAN M.W.;
 CC RT "Rabbit aortic smooth muscle cells express inducible macrophage
 CC scavenger receptor messenger RNA that is absent from endothelial

FT cells".
 RL J. Clin. Invest. 90:1450-1457(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93131972.
 RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
 RA MIYAZAKI T., NAKAMURA H., UESUGI S., IMANISHI T., KAMABE Y.,
 RA ITAKURA H., YAZAKI Y., MATSUMOTO A.;
 RT "Charged collagen structure mediates the recognition of negatively
 RT charged macromolecules by macrophage scavenger receptors.";
 RL J. Biol. Chem. 268:2126-2133(1993).
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEIN. IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L11693; AAA31402.1; -
 CC EMBL: L11692; AAA31403.1; -
 CC EMBL: D13381; BAA02649.1; -
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC DR PFAM: PF00530; SRCR; 1.
 CC DR PFAM: PF01391; Collagen; 1.
 CC KM Transmembrane; Glycoprotein; 1.
 CC FT DOMAIN 1 50
 CC FT TRANSMEM 1 73
 CC FT DOMAIN 74 74
 CC FT DOMAIN 74 109
 CC FT DOMAIN 110 272
 CC FT DOMAIN 273 344
 CC FT DOMAIN 345 454
 CC FT CARBOHYD 82 82
 CC FT CARBOHYD 102 102
 CC FT CARBOHYD 143 143
 CC FT CARBOHYD 184 184
 CC FT CARBOHYD 221 221
 CC FT CARBOHYD 249 249
 CC FT CARBOHYD 267 267
 CC FT VARSPLIC 348 354
 CC FT VARSPLIC 355 454
 CC FT CONFLICT 106 106
 CC SQ SEQUENCE 454 AA; 49745 MW; 5D780348 CRC32;
 CC -----
 CC Query Match 50.38; Score 381; DB 1; Length 454;
 CC Best Local Similarity 49.08; Pred. No. 5,89e-73;
 CC Matches 49; Conservative 20; Mismatches 31; Indels 0; Gaps 0;
 CC 353 VRLVGGSGPHEGVEIFHEGOWGTVCDDRWELRGLVCRSLGYKQGVSHKRAYGKGT 412
 CC 447 VRLMDENKKEGVEVFINGOWGTICDDGWTKDAVAIVRQLGYKGPAPARIMAYFGEGK 506
 CC DB 413 GPIWNEVFCFKESSIEECRIKRWGVYACSHSDAGVTC 452
 CC 507 GPIHVDNVKCTGNSRLADCIKODIGRHNCRHSDAGVIC 546
 CC -----
 CC RESULT 5
 CC ID MSRE_MOUSE STANDARD; PRT; 458 AA.
 CC AC 005585;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
 CC DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 CC DE LDL RECEPTOR I AND II).
 CC GN MSRL.
 CC OS Oryctolagus cuniculus (Rabbit).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 9301687.
 CC RA BICKEL P.E., FREEMAN M.W.;
 CC RT "Rabbit aortic smooth muscle cells express inducible macrophage
 CC scavenger receptor messenger RNA that is absent from endothelial


```

FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 240 349 CUB.
FT DOMAIN 357 519 MAM.
FT DOMAIN 539 649 CUB.
FT DOMAIN 656 694 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 693 786 SRCR.
FT ACT_SITE 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 986 986 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 236 BY SIMILARITY.
FT DISULFID 658 670 BY SIMILARITY.
FT DISULFID 665 683 BY SIMILARITY.
FT DISULFID 677 692 BY SIMILARITY.
FT DISULFID 787 911 INTERCHAIN (BY SIMILARITY).
FT DISULFID 825 841 BY SIMILARITY.
FT DISULFID 925 992 BY SIMILARITY.
FT DISULFID 956 971 BY SIMILARITY.
FT DISULFID 982 1010 BY SIMILARITY.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 170 170 POTENTIAL.
FT CARBOHYD 194 194 POTENTIAL.
FT CARBOHYD 283 283 POTENTIAL.
FT CARBOHYD 343 343 POTENTIAL.
FT CARBOHYD 350 350 POTENTIAL.
FT CARBOHYD 403 403 POTENTIAL.
FT CARBOHYD 455 455 POTENTIAL.
FT CARBOHYD 485 485 POTENTIAL.
FT CARBOHYD 518 518 POTENTIAL.
FT CARBOHYD 549 549 POTENTIAL.
FT CARBOHYD 645 645 POTENTIAL.
FT CARBOHYD 697 697 POTENTIAL.
FT CARBOHYD 701 701 POTENTIAL.
FT CARBOHYD 721 721 POTENTIAL.
FT CARBOHYD 740 740 POTENTIAL.
FT CARBOHYD 761 761 POTENTIAL.
FT CARBOHYD 804 804 POTENTIAL.
FT CARBOHYD 863 863 POTENTIAL.
FT CARBOHYD 902 902 POTENTIAL.
FT CARBOHYD 964 964 POTENTIAL.
SQ SEQUENCE 1034 AA; 114776 MW; 243B6471 CRC32;

ev Match 20.7% Score 157; DB 1; Length 1034;
st Local Similarity 31.7%; Pred. No. 2.57e-16;
lches 20; Conservative 18; Mismatches 23; Indels 2; Gaps 2;

```

```

RA CATERALL C.F., LYONS A., SIM R.M., DAY A.J., HARRIS T.J.R.;
RT "Characterization of primary amino acid sequence of human complement
RT control protein factor I from an analysis of cDNA clones.";
RN Biochem. J. 242:849-856(1987).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87280021.
RA GOLDBERGER G., BRUNS G.A.P., RITS M., EDGE M.D., KWATROWSKI D.J.;
RT "Human complement factor I: analysis of cDNA-derived primary
RT structure and assignment of its gene to chromosome 4.";
RN J. Biol. Chem. 262:10065-10071(1987).
CC -1- FUNCTION: FACTOR I IS RESPONSIBLE FOR CLEAVING THE ALPHA-CHAINS
CC OF C4B AND C3B IN THE PRESENCE OF THE COFACTORS C4-BINDING
CC PROTEIN AND FACTOR H RESPECTIVELY.
CC -1- SUBUNIT: HETERODIMER OF A LIGHT AND HEAVY CHAINS LINKED BY
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: PLASMA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
CC EMBL: Y00318; CAA68416.1; ALU_INIT.
CC EMBL: J02770; AAA52455.1; -.
CC PIR: A29154; A29154.
CC HSSP: P00763; LDPO.
CC SWISS-2DPAGE; P05156; HUMAN.
CC MIM: 217030; -.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS01209; LDLRA_1; 1.
CC PROSITE: PS0068; LDLRA_2; 2.
CC PFAM: PF00057; ldl_recept_a; 2.
CC PFAM: PF00089; trypsin; 1.
CC PFAM: PF00530; SRCR; 1.
CC Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 583 COMPLEMENT FACTOR I, HEAVY CHAIN.
FT CHAIN 19 335 COMPLEMENT FACTOR I LIGHT CHAIN.
FT CHAIN 340 583 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 220 258 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 257 295
FT DISULFID 229 247 BY SIMILARITY.
FT DISULFID 241 259 BY SIMILARITY.
FT DISULFID 259 271 BY SIMILARITY.
FT DISULFID 266 284 BY SIMILARITY.
FT DISULFID 278 293 BY SIMILARITY.
FT DISULFID 365 381 BY SIMILARITY.
FT DISULFID 467 531 BY SIMILARITY.
FT DISULFID 495 510 BY SIMILARITY.
FT DISULFID 521 550 BY SIMILARITY.
FT ACT_SITE 380 380 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 429 429 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 525 525 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 70 70 POTENTIAL.
FT CARBOHYD 103 103 POTENTIAL.
FT CARBOHYD 177 177 POTENTIAL.
FT CARBOHYD 464 464 POTENTIAL.
FT CARBOHYD 494 494 POTENTIAL.
FT CARBOHYD 536 536 POTENTIAL.
FT CONFLICT 558 558 V -> F (IN REF. 2).
SQ SEQUENCE 583 AA; 65720 MW; D673E9F0 CRC32;

Query Match 19.3%; Score 146; DB 1; Length 583;

```


RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RX SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE: 9404312.
RA LAVALLIE E.R., REHEMUTULA A., RACIE L.A., DIBLASIO E.A.,
RA FERENZ C., GRANT K.L., LIGHT A., MCCOY J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE-INTESTINE;
RC MEDLINE: 92189715.
RC LIGHT A., JANSKA H.;
RC "The amino-terminal sequence of the catalytic subunit of bovine
RC enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
EMBL: U09839; AAB40026.1; -
EMBL: L19663; AAL16035.1; -
PIR: A61436; A61436.
HSSP: P00763; IDPO.
DR PROSITE: PS00134; TRYPSIN_HTS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS50060; MAM_2; 1.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 2.
DR PFAM: PF00057; ldl_recept_a; 2.
DR PFAM: PF00089; trypsin; 1.
DR PFAM: PF00431; CUB; 2.
DR PFAM: PF00530; SRCR; 1.
DR PFAM: PF00629; MAM; 1.
DR PFAM: PF01390; SER; 1.
KM Signal-anchor: Glycoprotein; Myristate; Hydrolase;
KM Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
FT CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 240 350 CUB.
FT DOMAIN 358 520 MAM.
FT DOMAIN 540 650 CUB.

FT	DOMAIN	657	695	LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	657	787	SRCR.
FT	ACT_SITE	841	841	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	892	892	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	987	987	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYRISTATE (POTENTIAL).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	236	BY SIMILARITY.
FT	DISULFID	659	671	BY SIMILARITY.
FT	DISULFID	666	684	BY SIMILARITY.
FT	DISULFID	678	693	BY SIMILARITY.
FT	DISULFID	768	912	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	826	842	BY SIMILARITY.
FT	DISULFID	926	993	BY SIMILARITY.
FT	DISULFID	957	972	BY SIMILARITY.
FT	DISULFID	983	1011	BY SIMILARITY.
FT	CARBOHYD	116	116	POTENTIAL.
FT	CARBOHYD	147	147	POTENTIAL.
FT	CARBOHYD	170	170	POTENTIAL.
FT	CARBOHYD	194	194	POTENTIAL.
FT	CARBOHYD	233	233	POTENTIAL.
FT	CARBOHYD	263	263	POTENTIAL.
FT	CARBOHYD	264	264	POTENTIAL.
FT	CARBOHYD	404	404	POTENTIAL.
FT	CARBOHYD	456	456	POTENTIAL.
FT	CARBOHYD	486	486	POTENTIAL.
FT	CARBOHYD	519	519	POTENTIAL.
FT	CARBOHYD	550	550	POTENTIAL.
FT	CARBOHYD	646	646	POTENTIAL.
FT	CARBOHYD	698	698	POTENTIAL.
FT	CARBOHYD	722	722	POTENTIAL.
FT	CARBOHYD	741	741	POTENTIAL.
FT	CARBOHYD	762	762	POTENTIAL.
FT	CARBOHYD	864	864	POTENTIAL.
FT	CARBOHYD	903	903	POTENTIAL.
FT	CARBOHYD	965	965	POTENTIAL.
FT	VARSPLIC	166	192	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	808	808	R -> Y (IN REF. 3).
SO	SEQUENCE	1035 AA;	114887 MM;	5173034A CMC32.

Query Match 16.5%; Score 125; DB 1; Length 1035;
Best Local Similarity 30.6%; Pred. No. 2,51e-09;
Matches 19; Conservative 13; Mismatches 28; Indels 2; Gaps 2;

DB	694	VLFNCTTSSGLNFERIOSIHVACAEWNTQISDVQVLGL-GTGN-SVPIFSTGG 751
QY	447	VRLMGENKREGEVFEVINGMGITCDGMDKDAVIGRQDYGKGPARRATMAYFGEK 506
DB	752	GP 753
QY	507	GP 508

RESULT 14
ID ENTR HUMAN STANDARD; PRT: 1019 AA.
AC P98073;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 15-JUL-1998 (Rel. 36; Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DUDENUM;
RC MEDLINE: 95234679.
RA KITAMOTO Y., VEILE R.A., DONIS-KELLER H., SADLER J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinase,
RT the proteolytic activator of trypsinogen.";
RL Biochemistry 34:4562-4568(1995).

11

11

11

THIS PAGE BLANK (USPTO)